

AD-A235 652



DTIC

MAY 14 1991

c

D

2

LIKELIHOOD-RATIO DETECTION OF STOCHASTIC SIGNALS<sup>1,2</sup>

C.R. Baker\* and A.F. Gualtierotti\*\*

\*Department of Statistics  
University of North Carolina  
Chapel Hill, N.C. 27599-3260

\*\*IDHEAP  
University of Lausanne  
Lausanne, Switzerland 1015

LISS-40

October, 1989



Allocation For	
NTIS GRANT	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<input type="checkbox"/>
By _____	
Distribution _____	
Availability _____	
Dist	Availability
A-1	Special

<sup>1</sup>To appear, *Advances in Statistical Signal Processing. Volume 2: Signal Detection*. H.V. Poor and J.B. Thomas, eds., JAI Press.

<sup>2</sup>Preparation supported by ONR Grant N00014-89-J-1175 and NSF Grant NCR8713726.

DTIC COPY

91 5 13 066

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State and ZIP Code) University of North Carolina Chapel Hill, North Carolina 27514			7b. ADDRESS (City, State and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-89-J-1175		
8c. ADDRESS (City, State and ZIP Code) Statistics & Probability Program Arlington, VA 22217			10. SOURCE OF FUNDING NOS.		
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
11. TITLE (Include Security Classification) Likelihood-ratio Detection of Stochastic Signals			WORK UNIT NO.		
12. PERSONAL AUTHOR(S) C.R. Baker and A.F. Gualtierotti					
13a. TYPE OF REPORT TECHNICAL		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Yr., Mo., Day) 1989 Oct. (Rev. Feb. 1990)	
15. PAGE COUNT 49					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB. GR.			
			NonGaussian signal detection, optimum detection, stochastic signals.		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>Algorithms are given for the detection of a general stochastic signal imbedded in Gaussian or nonGaussian spherically-invariant noise. A summary is given of the theoretical development leading to the algorithms, along with the approximations leading to the algorithms. Adaptive and nonadaptive algorithms are described, along with specific procedures for their implementation.</p>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input checked="" type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION		
22a. NAME OF RESPONSIBLE INDIVIDUAL C.R. Baker			22b. TELEPHONE NUMBER (Include Area Code) (919) 962-2189		22c. OFFICE SYMBOL

## 1. INTRODUCTION

Detection of stochastic signals in Gaussian or nonGaussian noise is a valid model for many important signal detection problems. In some of these problems, the noise is very nonstationary and the signal cannot be represented as a set of narrowband components. Many examples can be given of applications where such problems arise; they abound in such areas as sonar and radar. In these situations, methods based on spectral analysis, assumptions of stationarity, approximations by matched filtering, etc. are all likely to be unsatisfactory.

Ideally, one would use signal detection algorithms based on the likelihood ratio. The main difficulty when the noise is Gaussian is that the finite-dimensional distributions of the signal-plus-noise process are typically unknown. This is especially true when the signal-plus-noise process is nonGaussian, so that its distribution cannot be characterized by its covariance and mean functions. This difficulty has led to the use of various suboptimum procedures, typically based on second-moment criteria.

However, for a significant class of problems involving stochastic signals in Gaussian noise, one can give a likelihood-ratio-based detection algorithm which does not require knowledge of the finite-dimensional distributions of the signal-plus-noise process. The development in descriptive fashion of such an algorithm (in two versions) is the principal content of this chapter.

In addition to discussion of signal detection for stochastic signals in Gaussian noise, a much shorter discussion will be given for detection in nonGaussian noise of a Gaussian mixture type. These results build upon those obtained for detection in Gaussian noise.

## 2. PRIOR WORK

Likelihood-ratio-based signal detection results for problems involving

Gaussian noise have largely involved a few special cases:

- (1) A known signal, for which the matched filter is a likelihood ratio test statistic [1], [2], and signals known except for phase and amplitude [1], [2].
- (2) A Gaussian signal-plus-noise process, for which quadratic-plus-linear operations are optimum [3], [4].
- (3) A possibly nonGaussian signal when the noise is a Wiener process [5]-[9]
- (4) A possibly nonGaussian signal in Gaussian noise when signal and noise are independent [6], [10].

Category (1) above is not of primary interest here.

Category (2) contains a class of problems that is often encountered in applications such as sonar. The implementation of a likelihood-ratio detection algorithm for discrete-time data requires knowledge of only the covariance matrices and mean vectors of the noise and signal-plus-noise processes. It is frequently possible to obtain good estimates of the noise covariance matrix; the noise is always present in typical applications, and sufficiently slowly-varying in its statistical properties to enable a good estimate to be made of its covariance matrix; the mean is frequently zero. We shall assume throughout this chapter that the noise covariance can be reliably obtained and that the noise has zero mean. It is quite a different matter for the signal-plus-noise covariance and mean. The signal is frequently present for only relatively short periods, and its properties may be changing too rapidly to enable one to obtain a reasonable estimate of its covariance and mean.

This problem of determining the covariance and mean of the signal-plus-noise process limits the usefulness of the likelihood ratio in many applications for which category (2) holds. However, even if one is able to

determine these quantities, the Gauss-vs.-Gauss model is too restricted to effectively model many real-world situations.

Categories 3 and 4 are of most interest in the present context. With regard to (3), actual use of a likelihood-ratio detection algorithm has been limited by the fact that conditions for its application are rarely satisfied. The Wiener process has very unusual properties, not encountered in applications such as sonar or radar. For example, a Wiener process has sample paths that are almost surely non-differentiable at every time point  $t$ ; also, the process is a martingale and a Markov process. These properties, even singly, are typically not found in applications. Algorithms based on the assumption of Wiener process noise are thus very unlikely to perform in a satisfactory fashion.

As for (4), conditions for existence of a likelihood ratio (nonsingular detection) have been given [6], [10]. The results of [10] also show how one can in principle determine the likelihood ratio when it exists. However, the knowledge required in order to actually carry this out will seldom be available, comprising as it does the finite-dimensional distributions of the signal.

Since signal-plus-noise probability distributions are so difficult to obtain, suboptimum procedures not requiring this knowledge have been extensively used. The deflection criterion gives a condition for optimality which dates back to at least the early 1940's [11]. The criterion requires one to specify a class of admissible test statistics,  $\mathcal{T}$ . The deflection of  $\tau$  in  $\mathcal{T}$  is defined by  $D(\tau) = [E_{S+N}\tau - E_N\tau]^2 / [E_N\tau^2 - (E_N\tau)^2]$ , where  $E_{S+N}(\cdot)$  and  $E_N(\cdot)$  denote expectation with respect to signal-plus-noise and noise, respectively. Deflection is used as the measure of performance. For the case where the operations on the data are assumed to be quadratic-plus-linear, and where the noise is Gaussian, solutions for the optimum operations under this

criterion were given in [12] and [13]. In this form, the solution requires only knowledge of the mean and the covariance of the signal-plus-noise process. If the signal-plus-noise process is also Gaussian, then the likelihood ratio yields a quadratic-plus-linear operation as a test statistic. However, this likelihood-ratio test statistic is not the same quadratic-plus-linear operation as that obtained using the deflection criterion, although there are some interesting relations between the two [13].

It should be noted that the Gaussian problem does not necessarily yield a bounded quadratic-linear operation for the infinite-dimensional problem [3]; similarly, the supremum of the deflection over all bounded quadratic-linear operations may not be achievable [13].

For a more detailed summary of results mentioned above, and additional references, see [14].

### 3. OPTIMUM DETECTION OF STOCHASTIC SIGNALS IN GAUSSIAN NOISE

#### 3.1. Introduction

Detection of stochastic signals imbedded in Gaussian or near-Gaussian noise is a common problem in such areas as sonar or radar. The prevalence arises because of physical properties of the medium and the signal source; see [15], [16] and the references given there for examples and discussion in sonar applications. The importance of the problem has been long-recognized; solutions have lagged behind. Of course, one model for which detection algorithms are well-known is that of a signal known except for phase and amplitude. This is not, however, a sufficiently rich class of signal processes to adequately model many of the important sonar-radar problems.

Ideally, one would have a detection algorithm having the following properties:

- (1) Likelihood-Ratio-Based. Monotone functions of the likelihood ratio are optimum test statistics under a variety of criteria [1].
- (2) Information-Preserving. Test statistics based on independent sampling of the data will in general destroy information contained in the continuous-time waveform. A continuous-time algorithm should be obtained, then approximated as closely as possible by a discrete-time algorithm for digital implementation.
- (3) Implementable. An abstract expression for the likelihood ratio, based on mathematical quantities that cannot be obtained in applications, is useless as a detection algorithm. The same comment applies to test statistics whose implementation depends upon an unrealistically-detailed knowledge of the physical environment.
- (4) Adaptive. The algorithm should have the capability of adapting to changes in the environment or the signal source.

Two algorithms will now be described. One meets all four of the criteria mentioned above. The second is not adaptive, and requires more prior information than the first, but is more powerful when it can be implemented.

The development will follow the path along which the algorithms were originally obtained. A fixed Gaussian noise ( $N_t$ ) is considered; one is attempting to detect a signal imbedded in this noise. The development begins by considering the general continuous-time problems.

First, a characterization is given of processes ( $Y_t$ ) for which the likelihood ratio of  $Y$  w.r.t.  $N$  exists. Conditions are then given that guarantee existence of a likelihood ratio. Next, assuming that these sufficient conditions are satisfied, representations of likelihood ratios are obtained. These results are all for the continuous-time problem. In current practice, detection algorithms will typically be implemented in digital format

after sampling. This requires that the continuous-time likelihood ratio be approximated, and procedures for implementation of a detection algorithm specified. This is the next step in the development, with a partially-recursive formulation. Two versions of the algorithm are given.

The development can be given in several forms and at various levels of complexity. We have elected to give the framework and results for the general continuous-time problem in precise mathematical language. However, the proofs are only summarized, with references to the original paper [17] for the detailed versions.

Some discussions and interpretations of the main results for the continuous-time problem are also included.

The development of the discrete-time approximation results in an implementable algorithm based upon the continuous-time likelihood ratio. Our objective in that section is to reach those readers who may have an interest in actually implementing a detection algorithm. Thus, this part can be read without serious reference to the development for the continuous-time analysis. However, the results on approximation and implementation cannot be fully appreciated without keeping in mind that they are derived from an analysis of the continuous-time problem that makes no significant limiting assumptions on the properties of either the noise or the signal-plus-noise processes, and that the discrete-time algorithm is obtained as an approximation to a continuous-time log-likelihood ratio.

The reader interested in genuine engineering applications can find discrete-time implementations of detection algorithms in Section 3.5.

For the continuous-time problem, the basic setup is as follows.  $(N_t)$  and  $(Y_t)$ ,  $t$  in  $[0,1]$ , are real-valued stochastic processes on  $(\Omega, \mathcal{F}, P)$ .  $(N_t)$  is Gaussian, m.s. continuous, separable, and vanishes a.s. (almost surely) at



$t = 0$ .  $\nu_Y$  and  $\nu_N$  are the induced measures on  $\mathbb{R}^{[0,1]}$ , the set of all real-valued functions on  $[0,1]$ .  $\mu_Y$  and  $\mu_N$  are the induced measures on  $L_2[0,1]$  (when  $(Y_t)$  has paths a.s. in  $\mathcal{L}_2[0,1]$ ).  $\mu_N$  is assumed to have infinite-dimensional support. The following problems are considered.

- (1) Determine conditions for existence of the likelihood ratios  $d\nu_Y/d\nu_N$  and  $d\mu_Y/d\mu_N$ ;
- (2) When absolute continuity holds, find  $d\nu_Y/d\nu_N$  and  $d\mu_Y/d\mu_N$ .

The interval  $[0,1]$  is selected only for convenience; the results hold for any finite interval.

### 3.2. Mathematical preliminaries

All stochastic processes will be real-valued and defined on a probability space  $(\Omega, \mathcal{F}, P)$  with index set  $[0,1]$ , unless otherwise noted. For a stochastic process  $(V_t)$ ,  $\sigma_t^0(V)$  will denote the  $\sigma$ -field generated by  $\{V_s, s \leq t\}$ , and  $\sigma_t(V)$  its  $P$ -completion.  $\underline{\sigma}^0(V)$  and  $\underline{\sigma}(V)$  are the corresponding filtrations:  $\underline{\sigma}(V) = \{\sigma_t^0(V), 0 \leq t \leq 1\}$ .  $\bar{L}_t(V)$  is the closed linear span in  $L_2[P]$  of  $\{V_s, s \leq t\}$ .

Let  $\underline{\sigma}(Y)$  be the completed filtration determined by a stochastic process  $(Y_t)$ . The predictable  $\sigma$ -field  $\mathcal{P}[\underline{\sigma}(Y)]$  is the smallest  $\sigma$ -field in  $\Omega \times [0,1]$  containing all sets of the form  $\{(\omega, t): X(\omega, t) \in A\}$ , where  $A$  is any Borel set in  $\mathbb{R}$  and  $(X_t)$  is any process adapted to  $\underline{\sigma}(Y)$  and having continuous (w.p. 1) paths. A stochastic process  $(W_t)$  is  $\underline{\sigma}(Y)$ -predictable if the map  $(\omega, t) \rightarrow W(\omega, t)$  is  $\mathcal{P}[\underline{\sigma}(Y)]/\mathcal{B}[\mathbb{R}]$  measurable.

$(N_t)$  will denote a Gaussian process that is m.s. (mean-square) continuous, separable with respect to closed sets, zero-mean, and vanishes (with probability one) at  $t = 0$ . We assume WLOG that  $\mathcal{F}$  is the smallest  $\sigma$ -field containing  $\underline{\sigma}(N)$ :  $\mathcal{F} = \sigma_1(N)$ .

For measures  $\tau$  and  $\gamma$  defined on the same  $\sigma$ -field  $\mathcal{A}$ ,  $\tau \ll \gamma$  ( $\tau$  absolutely continuous w.r.t.  $\gamma$ ) if for all sets  $A$  in  $\mathcal{A}$  such that if  $\gamma(A) = 0$ , one has also  $\tau(A) = 0$ .  $\tau \sim \gamma$  ( $\tau$  equivalent to  $\gamma$ ) if  $\gamma(A) = 0 \Leftrightarrow \tau(A) = 0$ .

For a positive integer  $M < \infty$ ,  $\mathcal{B}^M$  will be the Borel  $\sigma$ -field of  $\mathbb{R}^M$  under the product topology.  $C_0[0,1] \equiv C_0$  is the set of all real-valued functions that are continuous on  $[0,1]$  and vanish at  $t = 0$ .  $C_0$  is endowed with the sup norm topology, and  $\mathcal{C}$  is the resulting Borel  $\sigma$ -field, also generated by the evaluation maps  $\{\pi_t, 0 \leq t \leq 1\}$ ,  $\pi_t(x) = x_t$ . The Borel  $\sigma$ -field of  $C_0^M$  is  $\mathcal{C}^M$ , the product  $\sigma$ -field of  $M$  copies of  $\mathcal{C}$ .  $\mathbb{R}^{[0,1]}$  is the space of real-valued functions on  $[0,1]$  and  $\mathcal{B}^{[0,1]}$  is the  $\sigma$ -field in  $\mathbb{R}^{[0,1]}$  generated by the cylinder sets  $\{x: (x(t_1), \dots, x(t_n)) \in A^n\}$  for  $n \geq 1$ ,  $t_1, \dots, t_n$  in  $[0,1]$ , and  $A^n \in \mathcal{B}[\mathbb{R}^n]$ .

$E(\cdot)$  will denote mathematical expectation; the underlying measure will be clear from the context.

The approach used here makes extensive use of the spectral (Cramér-Hida) representation of  $(N_t)$  and its properties [18], [19].  $(N_t)$  has a proper canonical representation, which we assume to have the form

$$N_t(\omega) = \sum_{i=1}^M \int_0^t F_i(t,s) B_i(\omega, ds). \quad (2.1)$$

Without the assumption that  $(N_t)$  is Gaussian, the  $B_i$ 's are zero-mean orthogonal-increment processes, mutually orthogonal and m.s. continuous. Their nondecreasing variances  $EB_i^2(\cdot)$  define Borel measures  $\beta_i$  on  $[0,1]$  in the usual way:  $\beta_i(a,b] = EB_i^2(b) - EB_i^2(a)$ ; moreover,  $\beta_{i+1} \ll \beta_i$  for  $i \geq 1$ . Each  $F_i: [0,1] \times [0,1] \rightarrow \mathbb{R}$  is Borel-measurable,  $F_i(t,x) = 0$  for  $x > t$ , and

$$\sum_{i=1}^M \int_0^1 \int_0^1 F_i^2(t,s) d\beta_i(s) dt < \infty.$$

$M \leq \infty$  is the multiplicity of  $(N_t)$ .

The proper canonical representation (2.1) has the property that  $\bar{L}_t(N) = \bar{L}_t(B)$  for all  $t$  in  $[0,1]$ , and that  $\sum_{i=1}^M \int_0^1 F_i(t,s) g_i(s) d\beta_i(s) = 0$  for all  $t$  in  $[0,1]$  if and only if  $g_i = 0$  in  $L_2[\beta_i]$ , all  $i \leq M$ . Thus  $\{F_i(t, \cdot), t \in [0,1]\}$  spans  $L_2[\beta_i]$ .

The representation (2.1) is an equality in the mean-square sense, thus holds a.e.  $dP$  for each fixed  $t$ . However, taking both sides of (2.1) separable w.r.t. closed sets gives path equality a.e.  $dP$ .

The assumption that  $(N_t)$  is Gaussian further implies that the  $B_i$ 's are a Gaussian family; thus they are mutually independent, have independent increments, and hence can be assumed to be path-continuous. Moreover, equality of  $\bar{L}_t(N)$  and  $\bar{L}_t(B)$  implies that  $\underline{\sigma}_t(N) = \underline{\sigma}_t(B)$ , for all  $t$  in  $[0,1]$ . Each  $(B_i(t))$  is a martingale w.r.t.  $\underline{\sigma}(B)$ , from the independence of the  $B_i$ 's, and thus also w.r.t.  $\underline{\sigma}(N)$  and w.r.t.  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$ , where  $(V_t)$  is any stochastic process independent of  $(N_t)$ .

$M$  is the multiplicity of  $(N_t)$ . We assume throughout, unless otherwise noted, that  $M < \infty$ . An extension of the results given here to Gaussian processes of infinite multiplicity would require some results on infinite-dimensional stochastic calculus that are not readily available.

For a vector stochastic process  $(\underline{V}_t)$  having paths a.s. in  $C_0^M$ ,  $P_{\underline{V}}$  will denote the induced measure on  $C_0^M$ :  $P_{\underline{V}}(A) = P \circ \underline{V}^{-1}(A)$ , where  $\underline{V}$  is the path map. Similarly, if  $(V_t)$  is a scalar process, then  $\nu_V$  will be the probability induced on  $\mathcal{A}^{[0,1]}$  by the process. If  $(V_t)$  is measurable and has paths a.s. in  $\mathcal{L}_2[0,1]$  (Lebesgue square-integrable functions on  $[0,1]$ ) then  $\mu_V$  is the induced measure on the Borel  $\sigma$ -field of  $L_2[0,1]$  and  $V$  will denote the path map of  $\Omega$  into  $L_2[0,1]$ .

We assume WLOG that  $\text{support}(P_{\underline{B}}) = C_0^M$  and that  $\text{supp}(\mu_N) = L_2[0,1]$ . Since these measures are Gaussian, their supports are closed linear manifolds equal

to the closure of the ranges of their covariance operators [20]. One can thus always work with this subspace; it preserves the original linear space structure under the original norm (and inner product, for  $L_2[0,1]$ ).

The representation (2.1) and the usual properties of a m.s. continuous process generate a family of real separable Hilbert spaces. For a m.s. continuous process  $(V_t)$ , let  $H_V$  denote its RKHS (reproducing kernel Hilbert space), with inner product  $\langle \cdot, \cdot \rangle_V$ . Let  $r_V$  be the covariance function of  $(V_t)$  and  $R_V$  its (trace-class) covariance operator in  $L_2[0,1]$ ;  $R_V$  can be represented as an integral operator with  $r_V$  as its kernel. All elements in  $H_V$  are continuous functions on  $[0,1]$ .  $R_V$  defines a Hilbert space  $\hat{H}_V$  of  $L_2[0,1]$  elements, consisting of  $\text{range}(R_V^{1/2})$  together with the inner product

$$\langle h, g \rangle_{\hat{H}_V} = \sum_n \langle g, u_n \rangle \langle h, u_n \rangle / \rho_n$$

where  $\{\rho_n, n \geq 1\}$  and  $\{u_n, n \geq 1\}$  are the non-zero eigenvalues and associated orthonormal eigenvectors of  $R_V$ , and  $\langle \cdot, \cdot \rangle$  is the  $L_2[0,1]$  inner product. Since it is assumed that  $\text{support}(\mu_N) = L_2[0,1]$ , the eigenvalues  $\{\lambda_n, n \geq 1\}$  of  $R_N$  are all non-zero and the associated orthonormal eigenvectors are complete in  $L_2[0,1]$ . For  $L_2[\beta_i]$ ,  $i \leq M$ , already defined,  $H$  will denote  $\bigoplus_{i=1}^M L_2[\beta_i]$ : functions  $\underline{f}$  of the form  $\underline{f} = (f_1, f_2, \dots, f_M)$  with each  $f_i$  in  $L_2[\beta_i]$ , and with inner product of  $\underline{f}$  and  $\underline{g}$  given by  $\sum_{i=1}^M \int_0^1 f_i(s) g_i(s) d\beta_i(s)$ . Unitary maps between these Hilbert spaces are described in the following lemma.

**Lemma 3.1.** Define the following linear maps.

$$U_1: H_N \rightarrow \hat{H}_N, \quad U_1 g = [g].$$

where  $[g]$  is the equivalence class in  $L_2[0,1]$  generated by  $g$ .

$$U_2: H \rightarrow H_N, \quad [U_2 q](t) = \sum_{i=1}^M \int_0^t F_i(t,s) g_i(s) d\beta_i(s),$$

where  $q = (g_1, \dots, g_M)$ .

$$U_3: \bar{L}_1(N) \rightarrow H_N, \quad U_3 N_t = R_N(t, \cdot).$$

$$U_4: H \rightarrow H_B, \quad (U_4 q)_i(t) = \int_0^t g_i(s) d\beta_i(s).$$

Then: the maps  $U_1$ ,  $U_2$ ,  $U_3$ , and  $U_4$  are unitary.

Corollary. Let  $f_t$  be the vector function with  $i$ th component  $f_t^i(s) = \int_0^t F_i(s,u) d\beta_i(u)$ . Then  $f_t^i$  is in  $H_N$  and  $\langle f_t^i, f_s^j \rangle_N = \delta_{ij} EB_i(t) B_j(s)$  for all  $i, j \leq M$  and  $s, t$  in  $[0,1]$ .

For the main results to be discussed below, a sketch of each proof will be given. For the sake of illustration, we will assume in each sketch that  $(N_t)$  has multiplicity  $M = 1$  and that  $B_1 = B$  is the standard Wiener process,  $EB^2(\cdot, t) = t$ . Thus,  $N(t) = \int_0^t F(t,s) dB(s)$ . This assumption of unit multiplicity and  $B_1$  the Wiener process is only used in the sketches to simplify the illustration. The detailed proofs are in [17].

Theorem 3.1. For  $i \leq M$ , let  $(\phi_i(t))$  be a measurable stochastic process with paths a.s. in  $\mathcal{L}_2[\beta_i]$ . Define a vector process  $(Z_t)$  with paths a.s. in  $C_0^M$  by

$$Z_i(t) = \int_0^t \phi_i(x) d\beta_i(x) + B_i(t)$$

and stochastic processes  $(S_t)$  and  $(Y_t)$  by

$$S_t = \sum_{i=1}^M \int_0^t F_i(t,s) \phi_i(s) d\beta_i(s), \quad Y_t = \sum_{i=1}^M \int_0^t F_i(t,s) dZ_i(s).$$

Then there exists  $\Phi: C_0^M \rightarrow \mathbb{R}^{[0,1]}$  that is  $\mathcal{G}/\mathcal{A}^{[0,1]}$  measurable, and such that

- (1) For all  $t$  in  $[0,1]$ ,  $N_t(\omega) = \pi_t(\Phi[B(\omega)])$  a.s.  $dP(\omega)$ ;
- (2) If  $P_Z \ll P_B$ , then for all  $t$  in  $[0,1]$ ,  $Y_t(\omega) = \pi_t(\Phi[Z(\omega)])$  a.s.  $dP(\omega)$ .

Moreover, there exists a map  $\Phi_1: C_0^M \rightarrow L_2[0,1]$  such that  $\Phi_1$  is Borel-measurable, and

- (1<sub>1</sub>)  $N(\omega) = \Phi_1[B(\omega)]$  a.e.  $dP(\omega)$ ;
- (2<sub>1</sub>) If  $P_Z \ll P_B$ , then  $X(\omega) = \Phi_1[Z(\omega)]$  a.e.  $dP(\omega)$ .

If  $(N_t)$  has continuous paths, then there exists a map  $\Phi_2: C_0^M \rightarrow C_0$  which is  $\mathcal{G}/\mathcal{G}$  measurable, and

- (1<sub>2</sub>)  $N(\omega) = \Phi_2[B(\omega)]$  a.e.  $dP(\omega)$ ;
- (2<sub>2</sub>) If  $P_Z \ll P_B$ , then  $Y(\omega) = \Phi_2[Z(\omega)]$  a.e.  $dP(\omega)$ .

None of the statements of this theorem require that  $(N_t)$  be Gaussian, nor that the multiplicity  $M$  be finite.

Remarks. As noted, the results of Theorem 3.1 do not require that  $(N_t)$  be Gaussian. However, when  $(N_t)$  is Gaussian, and one further assumes that  $(\phi_i)$  is  $\mathcal{G}(B)$ -adapted for  $i \leq M$ , then  $P_Z \ll P_B$  ([7], Theorem 7.2). Thus, the relations given in (2), (2<sub>1</sub>), and (2<sub>2</sub>) all hold in this case.

Sketch of Proof. Since  $EN_t^2 = \|F(t, \cdot)\|_{L_2[0,1]}^2$ , the map  $t \rightarrow F(t, \cdot)$  is uniformly continuous as a map from  $[0,1]$  into  $L_2[0,1]$ .  $F(\frac{k}{2^n}, \cdot)$  can thus be approximated by simple functions and  $F(t, \cdot)$  by functions equal to  $F(\frac{k}{2^n}, \cdot)$  on the interval  $I_{n,k} \equiv [\frac{k-1}{2^n}, \frac{k}{2^n}]$ , for  $k \leq 2^n$ .  $F(t, \cdot)$  can thus be approximated by sets of simple functions.

For every simple function  $s(x) = \sum_{i=1}^m c_i I_{[a_i, b_i]}(x)$  and every real-valued function  $f: [0,1] \rightarrow \mathbb{R}$ , one can define a functional

$$Q_s: f \rightarrow \sum_{i=1}^m c_i [f(b_i) - f(a_i)].$$

A map  $\Phi_n: C_0[0,1] \rightarrow \mathbb{R}^{[0,1]}$ , induced by  $F$ , is then defined as follows: on  $I_{n,k}$ ,  $\Phi_n[f] \equiv \gamma_n(f)$ , where  $\gamma_n$  is the function  $Q_s$  when the simple function  $s$  is an approximation to  $F(\frac{k}{2^n}, \cdot)$ .

The second-order properties of  $B$  are then sufficient to ensure that  $(\Phi_n)$  converges in probability with respect to  $P_B$ , with the limit being a measurable map  $\Phi$ .  $\Phi$  is then shown to have the desired properties.  $\square$

Theorem 3.1 is a key tool for obtaining the results to follow. Formally, of course, one can write  $N$  as the result of a mapping  $F$  on the space of continuous functions such that  $N = FB$ . Intuitively this makes sense: in the proper canonical Cramér-Hida representation,  $\sigma_t(N) = \sigma_t(B)$  for all  $t$  in  $[0,1]$ . Theorem 3.1 gives both existence and precision to this heuristic notion. It should perhaps be noted that the mapping  $\Phi: C_0^M \rightarrow \mathbb{R}^{[0,1]}$  gives a version of  $(N_t)$  from  $(B_t)$ ; similarly,  $\Phi$  gives a version of  $(X_t)$  from  $(Z_t)$ .

### 3.3. Absolute Continuity (Existence of the Likelihood Ratio)

In this section several results are given for existence of a likelihood ratio. If one has a process  $(Y_t)$ , interpreted as signal-plus-noise, then a likelihood ratio will exist if the measure induced by  $(Y_t)$  is absolutely continuous with respect to the measure induced by  $(N_t)$ . In signal detection terminology, this means that if the probability of false alarm of any test statistic is equal to zero, then the probability of detection for that test statistic is also equal to zero.

We will first give conditions that are necessary for existence of a likelihood ratio. These conditions characterize the processes  $(Y_t)$  such that  $dv_Y/dv_N$  and  $d\mu_Y/d\mu_N$  exist.

Theorem 3.2. Let  $(V_t)$  be a stochastic process independent of  $(N_t)$ .

Suppose that  $(Y_t)$  is a process such that  $v_Y \ll v_N$ .

If  $(Y_t)$  is adapted to  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$ , then  $Y_t = S_t + N_t^*$  a.e.  $dP$  for each fixed  $t$  in  $[0,1]$ , where  $(N_t^*)$  has the same finite dimensional distributions as  $(N_t)$ , and is adapted to  $\underline{\sigma}(Y)$ .  $N_t^* = \sum_{i=1}^M \int_0^t F_i(t,s) dB_i^*(s)$  a.e.  $dP$ , each fixed  $t$  in  $[0,1]$ , where the  $B_i^*$ 's are mutually independent zero-mean Gaussian process,  $(\underline{B}_t^*)$  has the same law as  $(\underline{B}_t)$ , and  $\underline{\sigma}(\underline{B}^*) = \underline{\sigma}(N^*)$ . Moreover,

$$S_t = \sum_{i=1}^M \int_0^t F_i(t,s) \phi_i(s) d\beta_i(s),$$

where  $(\phi_i(t))$ ,  $i \leq M$ , is a stochastic process that is  $\underline{\sigma}(Y)$ -predictable and has paths a.s. in  $\mathcal{L}_2[\beta_i]$ .

Remark. Use of the filtration  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$  permits application of these results to problems in several areas involving feedback. For example, in information theory, the "message"  $(V_t)$  can be a stochastic process independent of the channel noise  $(N_t)$ ; the transmitted signal depends in a causal manner on this message and the channel output, which is of the form  $Y_t = S_t + N_t$ , where  $S_t = f[V_s, Y_s, s \leq t]$  with  $f$  a "coding" function. Thus, one needs that  $(S_t)$  be adapted to  $\underline{\sigma}(V) \vee \underline{\sigma}(N)$ . In the present context, this filtration is important for signal detection purposes. The message process  $(V_t)$  is usually independent of the ambient noise, although the eventual received signal process may depend on both  $(V_t)$  and  $(N_t)$ . For example, in active sonar problems where reverberation is the dominant noise component, the transmitted



waveform is affected by the reverberation-causing scatterers between the transmitter and the target as it travels through the water, as is the eventual reflection from the target as this reflected waveform travels toward the receiver. Since  $(Y_t)$  will eventually represent signal-plus-noise, it is obviously important that  $(Y_t)$  be adapted to  $\underline{g}(V) \vee \underline{g}(N)$ .

Sketch of Proof. Let  $D$  denote the likelihood ratio (Radon-Nikodym derivative) of  $\nu_Y$  with respect to  $\nu_N$ , evaluated at the path  $N(\omega) = \{N(\omega, t), t \in [0, 1]\}$ . The conditional expectation  $E[D | \sigma_t(N)]$  has a representation as a stochastic integral with respect to  $B$ , since, by the proper canonical Cramér-Hida representation,  $\underline{g}_t(N) = \underline{g}_t(B)$ . Let  $g$  be the function resulting from this integral representation. Using Girsanov's theorem, one obtains a Brownian motion  $G$ , in terms of  $B$  and  $g$ , with respect to the probability  $Q$  defined by  $dQ = DdP$ . To represent  $Y$  as a signal imbedded in additive noise, one must express  $G$ ,  $B$ , and  $g$  as  $N$ ,  $Y$ , and  $s$ ,  $s$  being the "derivative" of the signal. This is done by using the predictability properties of  $g$ .  $\square$

Theorem 3.2 states that when there is absolute continuity for  $Y$  with respect to  $N$ , then  $Y$  has a signal-plus-noise representation,  $Y = S + N^*$ . It should be carefully noted that  $N^*$  is a process with the same family of finite-dimensional probability distributions as  $N$ , therefore the same law on  $\mathfrak{A}^{[0,1]}$ . However,  $(N_t^*)$  is adapted to  $\underline{g}(Y)$ . This is a fact of major importance. Roughly, it states that  $Y$  has a signal-plus-noise representation where the noise is a measurable function of the original noise and of the message process (represented by  $\underline{g}(V)$  in Theorem 3.2). However, since the likelihood ratio of  $Y$  w.r.t.  $N$  depends only on laws induced by the two processes, and not on the measurability (adaptation properties), the fact that  $N^*$  is a function of message and original noise does not affect the likelihood ratio. In sum,

the likelihood ratio of  $Y$  with respect to  $N^*$  is the same as that of  $Y$  with respect to  $N$ . This fact will be used later to partially justify the form of the likelihood ratio used in the discrete-time approximation.

Theorem 3.2 contains necessary conditions for existence of a likelihood ratio. We now turn to sufficient conditions.

Theorem 3.3. Let  $(V_t)$  be a stochastic process independent of  $(N_t)$ .

Suppose that  $(S_t)$  is a stochastic process adapted to  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$  and with paths a.s. in  $H_N$ .

- (1) If  $Y_t = S_t + N_t$  a.e.  $dP$ , for each fixed  $t$  in  $[0,1]$ , then  $\nu_Y \ll \nu_N$ .
- (2) If  $Y_t = S_t + N_t$  a.e.  $dtdP$ , then  $\mu_Y \ll \mu_N$ .

Sketch of Proof. Since the signal has paths which are almost surely in the reproducing kernel Hilbert space of the noise, for almost every  $\omega$ , there is (with respect to Lebesgue measure) an equivalence class of functions  $[s_\omega]$  in  $L_2[0,1]$  such that

$$S(\omega, t) = \int_0^t F(t, x)[s_\omega](x) dx.$$

The main step in the proof consists in showing that one can choose a representative  $s_\omega$  in each class  $[s_\omega]$  such that the resulting process  $(s_t)$ ,  $s(\omega, t) \equiv s_\omega(t)$ , is a predictable stochastic process, thus has adequate smoothness properties.

Once this is done, one turns to the well-known absolute continuity properties of translates of Wiener processes. Theorem 3.1 permits one to write  $\tilde{N} = \Phi[B]$ , where  $\tilde{N}$  is a version of  $Y$  and  $\Phi: C_0[0,1] \rightarrow \mathbb{R}^{[0,1]}$  is measurable. The above representation for  $(S_t)$  gives  $P_Z \ll P_B$ , where  $Z(\omega, t) = \int_0^t s(\omega, t) dt + B(\omega, t)$ , and by Theorem 3.1 again, one has a version  $\tilde{Y}$  of  $Y$  defined by  $\tilde{Y} = \Phi[Z]$ . Since  $\nu_{\tilde{Y}} = \nu_Y$  and  $\nu_{\tilde{N}} = \nu_N$ , the result follows.  $\square$

The results of Theorem 3.3 are those that will be used to obtain a detection algorithm for applications. They are also conditions that one may hope to verify for actual physical problems. For example, they show that the likelihood ratio will exist if the signal process has all (with probability one) sample paths belonging to the RKHS of the noise, and if the signal process up to time  $t$  (all  $t$  in  $[0,1]$ ) can be obtained as a function of the signal and noise up to time  $t$ . These are conditions that one may expect to be satisfied. They are actually stronger than the sufficient conditions given in Theorem 3.3, since those conditions assume only that  $(S_t)$  is adapted to the completed filtration  $\underline{\sigma}(V) \vee \underline{\sigma}(N)$ , rather than the uncompleted filtration  $\underline{\sigma}^0(V) \vee \underline{\sigma}^0(N)$ . It is well-known [21] that when  $(S_t)$  is adapted to  $\underline{\sigma}^0(V) \vee \underline{\sigma}^0(N)$ , then for each  $t$ ,  $S_t$  can be expressed (with probability one) as  $S(\omega, t) = G_t[V(\omega, \cdot), N(\omega, \cdot)]$ , where  $G_t$  is a Borel-measurable map on  $\mathbb{R}^{[0,1]}$  that acts only on  $V(\omega, \cdot)$  and  $N(\omega, \cdot)$  up to time  $t$ .

We have been separately considering conditions for existence of likelihood ratios for measures induced on  $\mathbb{R}^{[0,1]}$  and for measures induced on  $L_2[0,1]$ . In addition, if  $(N_t)$  has continuous paths, one may wish to consider existence of the likelihood ratio for the measures induced on the space of continuous functions. In the next result, relations are drawn for the existence of these several likelihood ratios.

Theorem 3.4. (1) Suppose that  $(Y_t)$  is separable for closed sets and adapted to  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$ . Then  $\nu_Y \ll \nu_N \Rightarrow (Y_t)$  a.s. in  $\mathcal{L}_2[0,1]$  and  $\mu_Y \ll \mu_N$ . The converse is false.

(2) Suppose that  $(N_t)$  has continuous paths, and let  $(Y_t)$  be any process with continuous paths. Define  $\lambda_Y$  and  $\lambda_N$  to be the induced measures on  $C$ . Then

$$\nu_Y \ll \nu_N \Leftrightarrow \lambda_Y \ll \lambda_N \Leftrightarrow \mu_Y \ll \mu_N.$$

$$\nu_N \ll \nu_Y \Leftrightarrow \lambda_N \ll \lambda_Y \Leftrightarrow \mu_N \ll \mu_Y.$$

These equivalences do not require that  $(N_t)$  be Gaussian or that  $N_0 = 0$  a.s. dP.

Sketch of Proof. (1) The representation of  $Y$  as a signal embedded in additive noise given in Theorem 3.2 holds almost surely at every fixed time point. The separability assumption then produces pathwise equality, and consequently, since the signal is continuous and the noise mean-square continuous, the observation has paths in  $\mathcal{L}_2[0,1]$ . One then applies Theorem 3.3 to obtain the first statement; the converse is proved by using an example involving a known signal.

(2) The natural injection map  $i_0: C_0[0,1] \rightarrow \mathbb{R}^{[0,1]}$  is  $\mathcal{C}/\mathcal{B}$  measurable, and this is enough to yield those assertions which pertain to absolute continuity for the spaces  $\mathbb{R}^{[0,1]}$  and  $C_0[0,1]$ . Those regarding  $C_0[0,1]$  and  $L_2[0,1]$  follow from a theorem due to Kuratowski which states that Borel-measurable 1:1 maps between complete separable metric spaces carry Borel sets into Borel sets [22], and the observation that  $C_0[0,1]$  can be continuously injected into  $L_2[0,1]$ . □

The following result illustrates some of the subtleties contained in the statements of Theorem 3.2. In particular, this result shows that the sufficient conditions of Theorem 3.3 are not also necessary for absolute continuity.

Theorem 3.5. Let  $(N_t)$  be any m.s. continuous zero-mean Gaussian process on  $[0,1]$ . Suppose that  $(Y_t)$  is a process adapted to  $\underline{\sigma}(V) \vee \underline{\sigma}(N)$ , where  $(V_t)$  is independent of  $(N_t)$ . If  $\nu_Y \ll \nu_N$  (resp.,  $\mu_Y \ll \mu_N$ ), then it is

not necessary that  $Y_t = S_t + N_t$  a.e.  $dP$  for each fixed  $t$  in  $[0,1]$ , where  $(S_t)$  is adapted to  $\underline{\sigma}(V) \vee \underline{\sigma}(N)$  and has almost all paths in  $H_N$  (resp.,  $\hat{H}_N$ ). If  $(N_t)$  and  $(Y_t)$  have continuous paths, then such a representation is not necessary for  $\lambda_Y \ll \lambda_N$ .

Sketch of Proof. Let  $Y$  be equal to  $V + N$ ,  $V$  Gaussian, independent of  $N$ . It is possible to choose  $V$  such that absolute continuity obtains while the paths of  $V$  are not (with probability one) in the reproducing kernel of the noise: the covariance of  $V$  yields an integral operator of the form  $R_N^{1/2} T R_N^{1/2}$  with  $T$  Hilbert-Schmidt, but not trace-class.  $\square$

#### 3.4. Likelihood Ratios

In this section, general expressions for likelihood ratios are obtained under the assumption that the sufficient conditions of Theorem 3.3 are satisfied. First, a result on the (regular) conditional probability of  $\underline{B}$  given  $N$  is obtained when  $N$  is viewed as a map into  $\mathbb{R}^{[0,1]}$ . Such a conditional probability exists, since  $(\underline{B}_t)$  takes its path values in the complete separable metric space  $C_0^M$  while  $(N_t)$  is defined on the same underlying probability space  $(\Omega, \mathcal{F}, P)$  as  $(\underline{B}_t)$  [23].

Let  $T_n = \{t_1^n, \dots, t_{p(n)}^n\} \subseteq ]0,1]$  be a strictly ordered set of  $p[n]$  elements, for every  $n$  in  $\mathbb{N}$ , and  $T = \bigcup_{n=1}^{\infty} T_n$ . The sets  $T_n$  are increasing and chosen so that  $T$  is dense in  $[0,1]$ . If  $S$  is a finite or countable set in  $[0,1]$  whose elements are ordered as  $\{s_1, s_2, s_3, \dots\}$  and  $f$  is a function defined on  $[0,1]$ , then  $w_S$  is the map defined by the relation

$$w_S f = (f(s_1), f(s_2), f(s_3), \dots).$$

$e_n$  is the map defined on  $\mathbb{R}^{\infty}$  which retains the first  $n$  components of every sequence.  $I_t^i$  will be the element of  $\bigoplus_{j=1}^M L_2[\beta_j]$  having  $i$ th component given by

$I_{[0,t]}$  and with all other components zero. Finally,  $R_{N,n}$  is the covariance matrix of the Gaussian vector  $w_{T_n} N$ .

In the course of this development, we obtain "inverses" to the maps  $\Phi$ ,  $\Phi_1$ , and  $\Phi_2$  defined in Theorem 3.1. These inverses are crucial to defining likelihood ratios.

Theorem 3.6. The conditional law  $Q(f, \cdot)$  of  $\underline{B}$  given that  $N = f$  is given by the relation

$$Q(f, C) = I_C(\underline{m}[f, \cdot]), \quad \nu_N - a.e. f,$$

where  $\underline{m}$  is a continuous Gaussian stochastic process defined on  $(\mathbb{R}^{[0,1]}, \mathcal{A}^{[0,1]}, \nu_N)$  which has the same law as  $\underline{B}$  with respect to  $P$ .  $\underline{m}$  is the weak limit of a sequence  $\{\underline{m}_n, n \in \mathbb{N}\}$  of continuous Gaussian stochastic processes defined on  $(\mathbb{R}^{[0,1]}, \mathcal{A}^{[0,1]}, \nu_N)$  which have components  $m_i^n$  given by the relation

$$m_i^n[f, t] = \langle w_{T_n} U_2 \{I_t^i\}, R_{N,n}^{-1} (e_{p[n]} \circ w_T)' \rangle_{\mathbb{R}^{p[n]}}.$$

$1 \leq i \leq M$  ( $U_2$  is defined in Lemma 3.1). Moreover,  $[\underline{m} \circ \Phi](\underline{c}) = \underline{c}$  a.e.  $dP_{\underline{B}}(c)$ .

Sketch of Proof. One first notices that it suffices to compute the conditional law of  $\underline{B}$  with respect to  $\underline{N}$ , where  $\underline{B}$  and  $\underline{N}$  are vectors obtained by sampling  $B$  and  $N$  at a dense and countable set of time points. The problem is further reduced to a finite-dimensional one by restricting attention to cylinder sets. One then uses the usual formulae for conditional laws of Gaussian random variables, and then attempts a limiting procedure.

The limiting procedure succeeds for the following reasons. Let  $\underline{B}_n$  and  $\underline{N}_n$  be the finite-dimensional Gaussian vectors obtained in the reduction outlined

above. The conditional law of  $\underline{B}_n$  given  $\underline{N}_p$  has a mean  $\underline{m}_{n,p}$  which depends on  $\underline{N}_p$ , and a covariance  $\underline{\Sigma}_{n,p}$  which does not. One must then observe that the elements of the mean and the covariance can be expressed in terms of finite-dimensional projections in the reproducing kernel Hilbert space of the noise  $N$ . The time points being dense and  $N$  being mean-square continuous ensure that these projections converge to the identity. The mean  $\underline{m}_{n,p}$  produces a continuous Gaussian process which has a weak limit  $m$ , with respect to  $P_N$ , and  $m$  is the point at which the mass of the conditional law of  $B$  given  $N$  is concentrated. Finally, the fact that  $m \circ \Phi(c) = c$  a.e.  $dP_B(c)$  follows from martingale equalities. Since this is not contained in [17], its proof will now be given in full.

By Theorem 3.1, with  $\pi_i[\underline{c}](t) \equiv c_i(t)$ , one has

$$\begin{aligned} P_{\underline{B}}\{\underline{c} \text{ in } C_0^M[0,1]: \sup_{0 \leq t \leq 1} |\pi_i[(m \circ \Phi)(\underline{c})](t) - \pi_i[\underline{c}](t)| > \epsilon\} \\ = P\{\omega: \sup_{0 \leq t \leq 1} |(m_1 \circ \tilde{N}[\omega])(t) - B_1(\omega, t)| > \epsilon\}, \end{aligned}$$

where  $(\tilde{N}_t)$  is a version of  $(N_t)$ ,  $\tilde{N}_t(\omega) = N_t(\omega) = \pi_t \Phi[B(\omega)]$  a.e.  $dP(\omega)$  for each fixed  $t$ . By Theorem 3.6,  $\underline{m}$  has w.r.t.  $\nu_N$  the same law as  $\underline{B}$ , so that  $((m \circ \tilde{N}[\omega])(t) - B(\omega, t))$  is a continuous square-integrable martingale. This gives

$$\begin{aligned} P\{\omega: \sup_{0 \leq t \leq 1} |(m_1 \circ \tilde{N}[\omega])(t) - B_1(\omega, t)| > \epsilon\} \\ \leq \frac{1}{\epsilon^2} E |(m_1 \circ \tilde{N})(1) - B_1(1)|^2 \\ = \frac{1}{\epsilon^2} \left[ E(m_1 \circ \tilde{N})^2(1) + E B_1^2(1) - 2E(m_1 \circ \tilde{N})(1) B_1(1) \right]. \end{aligned}$$

It now suffices to show that  $E(m_1 \circ \tilde{N})(1) B_1(1) = \beta_1^2(1)$ .  $m_1 \circ \tilde{N}(1)$  is the limit in  $L_2[dP]$  of  $m_1^n \circ \tilde{N}(1)$ , where  $[m_1^n \circ x](1) = \langle w_{T_n} f_1^1, e_{p[n]x} \rangle_N$ . Thus,

$$E(m_i \circ \tilde{N})(1)B_i(1) = \lim_n E(m_i^n \circ \tilde{N})(1)B_i(1)$$

$$= \lim_n \sum_{k,j=1}^{p[n]} R_{N,n}^{-1}(j,k) \int_0^{t_j^n} F_1(t_j^n, s) d\beta_1(s) \int_0^{t_k^n} F(t_k^n, v) d\beta_1(v)$$

$$= \|f_1^1\|_N^2 = \beta_1^2(1). \quad \square$$

We now go to the final step in the general continuous-time detection problem: specification of the likelihood ratio when absolute continuity exists.

Theorem 3.7. Suppose that  $(V_t)$  is a stochastic process independent of  $(N_t)$ , and that  $(S_t)$  is a stochastic process adapted to  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$  with almost all paths in  $H_N$ . Then  $S_t = \sum_{i=1}^M \int_0^t F_i(t,s) \phi_i(s) d\beta_i(s)$ , where  $(\phi_i(s))$  is  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$ -predictable and has paths a.s. in  $\mathcal{L}_2[\beta_i]$ . Define  $(\underline{Z}_t)$  to be the  $C_0^M$  process with  $i$ th component given by  $Z_i(t, \omega) = \int_0^t \phi_i(s, \omega) d\beta_i(s) + B_i(t, \omega)$ .  $P_{\underline{Z}} \ll P_{\underline{B}}$  and  $S_t + N_t = \sum_{i=1}^M \int_0^t F_i(t,s) dZ_i(s)$ . Then

(1) Suppose  $Y_t = S_t + N_t$  a.e.  $dP$ , for each fixed  $t$  in  $[0,1]$ . Then  $[d\nu_Y/d\nu_N](y) = [dP_{\underline{Z}}/dP_{\underline{B}}](\underline{m}y)$  a.e.  $d\nu_N(y)$ , where  $\underline{m}$  is defined as in Theorem 3.6.

(2) Suppose  $Y_t = S_t + N_t$  a.e.  $dtdP$ . Then  $[d\mu_Y/d\mu_N](x) = [dP_{\underline{Z}}/dP_{\underline{B}}](\underline{m}x)$  a.e.  $d\mu_N(x)$ , where  $\underline{m}(y)$  in  $C_0^M$  has as its  $i$ th component  $[\underline{m}(y)]_i(t) = \sum_{n \geq 1} \langle y, e_n \rangle \langle f_t^i, e_n \rangle / \lambda_n$ .

(3) Suppose that  $Y_t = S_t + N_t$  a.e.  $dP$  for each  $t$  in  $[0,1]$ , and that  $(N_t)$  and  $(Y_t)$  have continuous paths. Then  $[d\lambda_Y/d\lambda_N](x) = [dP_{\underline{Z}}/dP_{\underline{B}}](\underline{m}i_0x)$  a.e.  $d\lambda_N(x)$ , where  $\underline{m}$  is defined in (2) and  $i_0: C_0^M[0,1] \rightarrow L_2[0,1]$  is the injection map.



Sketch of Proof. This is a direct consequence of the fact that the maps  $\underline{m}$  defined in Theorem 3.6 and part (2) of Theorem 3.7 act as inverses of the maps  $\Phi$  and  $\Phi_1$ . The map  $m$  of part (1) was shown above to satisfy  $m \circ \Phi(c) = c$  a.e.  $dP_{\underline{B}}(c)$ . For the map  $m$  of part (2), the fact that  $m \circ \Phi_1(c) = c$  a.e.  $dP_{\underline{B}}(c)$  is proved by using the fact that  $P_{\underline{B}}$  is Gaussian and the definition of  $m$ .  $\square$

Corollary. Under the hypotheses of part (2) of Theorem 3.7,  $\underline{m}\Phi_1(x) = x$ , a.e.  $dP_{\underline{B}}(x)$ , with  $\Phi_1$  as defined in Theorem 3.1 and  $\underline{m}$  as defined in part (2) of Theorem 3.7.

In obtaining conditions for existence of the likelihood ratio, the maps  $\Phi$  and  $\Phi_1$  defined in Theorem 3.1 filled an essential role. Theorem 3.7 shows that their inverses (the  $\underline{m}$  maps) play a correspondingly-essential role in specifying the form of the likelihood ratios.

Explicit representations for the likelihood ratio  $dP_{\underline{Z}}/dP_{\underline{B}}$  can be obtained from known results, since  $(\underline{B}_t)$  has independent components that are path-continuous Gaussian martingales with respect to  $\underline{g}(N) \vee \underline{g}(V)$ . See, for example, [9]. It will be noted that although the proofs of absolute continuity require the proper canonical representation of  $(N_t)$ , the results do not require that this decomposition be known. This is a non-trivial consideration, since determining the decomposition is well-known to be a very difficult problem. However, the explicit expressions for the likelihood ratios do require knowledge of the proper canonical representation. This constitutes a significant problem in obtaining an implementable discrete-time detection algorithm based on the continuous-time likelihood ratio.

The first main objective has now been achieved: beginning with a Gaussian noise of a very general type, conditions have been obtained for existence of the likelihood ratio, and expressions have been obtained for the likelihood

ratio when it exists.

Thus, the first part of the program is complete. We now move to the second part, which is equally important: approximation and implementation of the likelihood ratio.

### 3.5. Discrete-Time Approximation and Implementations

The results given above are for continuous-time observations. The expressions for the likelihood ratio given in Theorem 3.7 are very general, but in that form are simply mathematical results. To be useful, they must be converted into signal detection algorithms. The goal is to obtain an algorithm that meets the four criteria discussed in Section 3.1:

- (1) Likelihood-ratio-based;
- (2) Information-preserving;
- (3) Implementable;
- (4) Adaptive.

Criteria (1) and (2) will be met by constructing two algorithms that are discrete-time approximations to the general expressions for the likelihood ratio obtained in Section 3.4 above. One of these algorithms, which will eventually be denoted as Version I, requires only knowledge of the noise covariance matrix; it is completely adaptive to the signal-plus-noise process. Moreover, it is easily implementable. Version I thus satisfies all four of the above criteria.

Version II algorithm requires prior knowledge of a two-variable function (the drift of a diffusion). If a model giving this function is not available, then the function can be estimated from a "training" ensemble of signal-plus-noise data. Thus it is not adaptive and is more difficult to implement, but can be expected to generally perform better than Version I when it can be implemented.

One further criterion will also be imposed on the implementation: a partially-recursive formulation.

We now proceed to the development of these two detection algorithms. Version I, fully adaptive, is based on the following additional assumptions:

- (A.1) The noise process has multiplicity  $M=1$ , and the process  $(B_1(t))$  is the standard Wiener process  $(W(t))$ ; thus  $N(t) = \int_0^t F(t,s)dW(s)$ , where  $F$  is a Volterra kernel with  $\int_0^1 \int_0^1 F^2(t,s)dsdt < \infty$ ;
- (A.2) The signal-plus-noise process can be represented as  $(Y(t) = \int_0^t F(t,s)dZ(s))$ , where the process  $(Z_t)$  is a diffusion with respect to a Wiener process  $\hat{W}$  and has memoryless drift function, so that

$$Z(t) = \int_0^t \theta[Z(s)]ds + \hat{W}(t), \quad (5.1)$$

where  $P\{\omega: \int_0^1 \theta^2[Z(\omega,s)]ds < \infty\} = 1$ .

The second algorithm, Version II, assumes (A.1) above and that the signal-plus-noise process can be represented as  $Y(t) = \int_0^t F(t,s)dZ(s)$ , where

$$(A.3) \quad Z(t) = \int_0^t \sigma[s, Z(s)]ds + \hat{W}(t), \quad (5.2)$$

where again  $\hat{W}$  is a Wiener process, and  $P\{\omega: \int_0^1 \sigma^2[s, Z(\omega,s)]ds < \infty\} = 1$ .

The assumption (A.1) is reasonable from several viewpoints, such as the fact that all stationary processes and all discrete-time processes are of unit multiplicity and that any Gaussian vector can be represented as the result of a lower-triangular matrix operating on white Gaussian noise. It is also known [24] that unit multiplicity processes are dense (in  $L_2[dPdt]$ ) in the class of m.s. continuous processes. One can show that the assumption (A.1) is satisfied whenever the noise process has a proper canonical representation

$N(t) = \int_0^t F(t,s)dB(s)$ , where the variance of  $(B(t))$  is an absolutely continuous function on  $[0,1]$ .

To clarify the significance of the assumptions (A.2) and (A.3), it is necessary to review well-known material on the representation of processes  $(Z_t)$  such that  $P_Z \ll P_W$ . From Theorem 7.11 of [9], any such  $(Z_t)$  must be a process of "diffusion type":

$$Z(\omega, t) = \int_0^t \gamma[s, Z(\omega, \cdot)]ds + \hat{W}(\omega, t) \quad (5.3)$$

where  $(\hat{W}_t)$  is a Wiener process adapted to  $\mathcal{G}(Z)$ , and  $\gamma$  is a function on  $[0,1] \times C_0[0,1]$  such that

- i) for all  $s$  in  $[0,1]$ ,  $\gamma(s, x)$  does not depend on  $x(t)$  for any  $t > s$ ;
- ii)  $P\{\omega: \int_0^1 \gamma^2[s, Z(\omega, \cdot)]ds < \infty\} = 1$ .

Using the results of Theorem 3.2 above, it can be seen that assumption (A.3) reduces to the assumption that the function  $\gamma$  is memoryless. Of course, there is a very large class of processes such that this assumption is satisfied. Note that this is emphatically not equivalent to the assumption that the observed signal-plus-noise process is the solution to a stochastic differential equation. To be precise, given that assumption (A.1) is satisfied, the assumption (A.3) states that the signal-plus-noise process can be represented as a filtered diffusion.

Assumption (A.3) is of course much weaker than (A.2); the latter assumes not only that the process  $Z$  defined above is a diffusion, but that the drift function is time-homogeneous:  $\gamma[s, Z(\omega, \cdot)] = \theta[Z(\omega, s)]$  a.e.  $dP(\omega)ds$ .

The reduction of the problem to the class of processes satisfying (A.2) is motivated by the goal of developing a likelihood-ratio-based detection algorithm that can operate without any prior knowledge of the signal

properties: a completely adaptive algorithm.

An algorithm will now be described, assuming (A.1) and (A.3), and uniform sampling of the continuous-time waveform. For the detection problem as defined above, applying Theorem 3.7 and known results for the Wiener process (see, e.g., [9] or [8]), the general form (under a mild restriction) of the likelihood ratio on  $L_2[0,1]$  is

$$[d\mu_{S+N}/d\mu_N](x) = \lim_n \exp [\Lambda^n(\delta_n(m[x]))]$$

where  $0 = t_0^n < t_1^n < t_2^n < \dots < t_n^n = T$  is a partition of  $[0,T]$  such that  $\sup_j |t_{j+1}^n - t_j^n| \rightarrow 0$ ,  $\delta_n(x) \equiv (x(t_1^n), x(t_2^n), \dots, x(t_n^n))$ , and

$$\begin{aligned} \Lambda^n[\delta_n[m(x)]] = & \sum_{i=0}^{n-1} \sigma(i, m[x](t_i^n))(m[x](t_{i+1}^n) - m[x](t_i^n)) \\ & - (1/2) \sum_{i=0}^{n-1} \sigma^2(i, m[x](t_i^n))(t_{i+1}^n - t_i^n). \end{aligned} \quad (5.4)$$

$m$  is defined (as  $\underline{m}$ ) in part (2) of Theorem 3.7, and the limit exists in the norm of  $L_1[\mu_N]$ .

It should be noted that this approximation does not arise from sampling of the continuous-time observation. The situation is much more complicated; the approximation is obtained by sampling of the continuous-time function  $m(x)$ , where  $x$  is the observation. The difficulty is that  $m$  will generally not be known.

The representation of  $(N(t))$  by  $N(t) = \int_0^t F(t,s)dW(s)$  yields that  $R_N = FF^*$ , where  $F$  is the integral operator with  $F(t,s)$  as its kernel, and  $F^*$  is its adjoint. This can be used to provide an approximation to the function  $m(x)$  appearing in (5.4) that does not require calculation of eigenvalues and eigenvectors.

First, notice that  $\langle e_j, f_t \rangle = \int_0^T \int_0^t F(s,u) du e_j(s) ds = \int_0^t \int_0^T F(s,u) e_j(s) ds du = [LF^* e_j](t)$ , where  $[Lf](t) \equiv \int_0^t f(v) dv$ .  $L$  is considered as a map from  $L_2[0,1]$  into  $C[0,1]$ . Using this, the expression for  $m$  given in (2) of Theorem 3.7 can be rewritten as

$$\begin{aligned} m[x](t) &= \lim_{k \rightarrow \infty} [LF^* \sum_{j=1}^k \langle e_j, x \rangle R_N^{-1} e_j](t) \\ &= \lim_{k \rightarrow \infty} [LF^* R_N^{-1} P_k x](t) \end{aligned} \quad (5.5)$$

where  $P_k x$  is the projection of the function  $x$  on the subspace in  $L_2[0,1]$  spanned by  $\{e_1, \dots, e_k\}$ . Since  $R_N^{-1} = F^{*-1} F^{-1}$ , the preceding becomes  $m[x](t) = \lim_{k \rightarrow \infty} [LF^{-1} P_k x](t)$ .

A basic difficulty is that (with probability one [10]) the observation  $x$  will not be in the domain of the operator  $F^{-1}$ , so that  $F^{-1}x$  is not defined. In fact,  $LF^{-1}$  will in general not be a bounded linear operator. However, for almost all sample functions  $x$  (either from noise or signal-plus-noise),  $m[x](\cdot)$  is a continuous function on  $[0,1]$ . Thus the map  $m$  is a linear operator from  $L_2[0,1]$  into  $C[0,1]$  whose domain includes (with probability one) all sample functions of the noise and signal-plus-noise processes.

The difficulty in implementation of the approximate likelihood ratio (5.4) will lie in determining the function  $\sigma$  and linear operator  $m$ .  $\sigma$  is a parameter of the signal-plus-noise process, and its estimation is a problem of considerable interest in stochastic processes (as the drift of a diffusion) and in stochastic filtering. The possibly unbounded linear operator  $m$ , mapping  $L_2[0,1]$  into  $C[0,1]$ , depends only on the covariance function of the noise. If  $\sigma$  is known or estimated, and if the noise covariance function is completely known, including its orthonormal eigenvectors and associated eigenvalues, then

the preceding expressions can be used to obtain a discrete-time finite-sample approximation to the likelihood ratio. Here we consider such approximations when one knows only the covariance matrix of the noise.

Let  $R_N^n$  denote the covariance matrix of the noise; one can write  $R_N^n = \Delta^n F_n F_n^*$ , where the matrix  $F_n$  is lower triangular and  $\Delta^n$  is the sampling interval (uniform sampling). Now, the expression for  $m$  given above is of the form

$$m[x](t) = \lim_{k \rightarrow \infty} [LF^{-1}P_k x](t),$$

where  $R_N = FF^*$ ,  $L$  is the integration operator, and  $P_k$  is the projection of  $x$  onto the subspace spanned by  $\{e_1, \dots, e_k\}$ , where  $\{e_n, n \geq 1\}$  are o.n.

eigenvectors of  $R_N$ . Thus, a reasonable procedure is simply to replace this expression by  $\underline{m}^n[\underline{x}^n] = \underline{L}_n F_n^{-1} \underline{x}^n$ , where  $\underline{x}^n$  is the observed data vector, an element of  $\mathbb{R}^n$ , and  $\underline{L}_n$  is the summation operator in  $\mathbb{R}^n$ :  $(\underline{L}_n \underline{x}^n)_j = \sum_{i=1}^j x_i$ .

However, without further analysis and justification this would simply be an *ad hoc* assumption. Thus, we now examine the relation of  $\underline{L}_n F_n^{-1} \underline{x}^n$  to  $m(x)$ , where  $\underline{x}^n = (x(\Delta^n), x(2\Delta^n), \dots, x(n\Delta^n))$ .

By the Corollary to Theorem 3.7,  $(m \circ \Phi_1)(x) = x$  a.e.  $dP_W(x)$ , where  $\Phi_1$  is defined in Theorem 3.1. Hence, the distribution of  $m(N)$  is given by  $P_W$ , so that the vector  $\delta_n[m(N)] = (m(N)(\Delta^n), m(N)(2\Delta^n), \dots, m(N)(n\Delta^n))$  has probability distribution  $P_W \circ \delta_n^{-1}$ , where  $\delta_n: C[0,1] \rightarrow \mathbb{R}^n$  is defined by  $\delta_n(x) \equiv (x(\Delta^n), x(2\Delta^n), \dots, x(n\Delta^n))$ . Similarly, defining  $\underline{m}^n(\underline{x}^n) = \underline{L}_n F_n^{-1} \underline{x}^n$ ,  $\underline{m}^n(\underline{N}^n)$  has probability distribution  $P_W \circ \delta_n^{-1}$ , from the definition of  $F_n$ . Thus, in the preceding expression (5.4) for  $\Lambda^n[\delta_n[m(x)]]$ , and setting  $t_i^n = i\Delta^n$ , one can replace  $\delta_n[m(x)]$  by  $\underline{m}^n[\underline{x}^n]$ ; with respect to  $\mu_N$ , the law of  $\Lambda^n[\underline{m}^n(\underline{x}^n)]$  will be the same as that of  $\Lambda^n[\delta_n[m(x)]]$ .

Next, suppose that the signal is present. In examining the relation between the law of  $\Lambda^n[\delta_n[m(x)]]$ , as given by (5.4), and the law of  $\Lambda^n[\underline{m}^n[\underline{x}^n]]$ , obtained by substituting  $\underline{m}^n(\underline{x}^n)$  for  $\delta_n[m(x)]$ , we make the following smoothness assumptions:

(a)  $\underline{F}_n(i,j) \cong F(i\Delta^n, j\Delta^n)$  for all  $i,j \leq n$ ;

(b) The law of the random vector in  $\mathbb{R}^n$  with  $(i+1)$  component given by

$\int_0^{i\Delta^n} \sigma[s, Z(s)] ds + \hat{W}([i+1]\Delta^n)$  is approximately the law of the random vector in  $\mathbb{R}^n$  with  $(i+1)$  component given by  $\Delta^n \sum_{k=1}^i \sigma[k\Delta^n, Z(k\Delta^n)] + \hat{W}([i+1]\Delta^n)$ , where

$$Z(t) = \int_0^t \sigma(s, Z(s)) ds + \hat{W}(t)$$

as in assumption (A.3).

Assumption (a) is essentially equivalent to assuming that

$$\int_0^{(i \wedge j)\Delta^n} F(i\Delta^n, s) F(j\Delta^n, s) ds \cong \Delta^n \sum_{k=1}^{i \wedge j} F(i\Delta^n, k\Delta^n) F(j\Delta^n, k\Delta^n)$$

for every  $i, j \leq n$ . It is thus an assumption on the smoothness of  $\{F(t, \cdot), t \text{ in } [0, 1]\}$ . Note, however, that the smoothness requirement applies, for fixed  $t$ , only to  $F(t, \cdot)$  restricted to  $[0, t]$ . Similarly, (b) amounts to a smoothness assumption on  $\sigma$ .

Under  $\mu_{S+N}$ , letting  $\underline{y}^n$  be the vector such that  $y^n(k) = \int_0^{k\Delta^n} F(k\Delta^n, s) dZ(s)$ ,

$$\underline{m}^n[\underline{x}^n] = \underline{L}_n \underline{F}_n^{-1} \underline{y}^n.$$

Assumptions (a) and (b) then give that the law of  $\underline{m}^n[\underline{x}^n]$  is approximately that of  $\delta_n[m(x)]$ , with the law of  $\delta_n[m(x)]$  being approximately that of the random vector with  $(i+1)$  component



$$\Delta^n \sum_{k=1}^i \sigma[k\Delta^n, Z(k\Delta^n)] + \hat{W}((i+1)\Delta^n).$$

Thus, with the smoothness assumptions (a) and (b),  $\Lambda^n[\delta_n[m(x)]]$  and  $\Lambda^n[\underline{m}^n[\underline{x}^n]]$  have approximately the same distribution w.r.t.  $\mu_{S+N}$ . From the nature of assumptions (a) and (b), it can be seen that (if  $F$  and  $\sigma$  are sufficiently smooth) the approximations can be expected to become better (for a fixed observation time) as  $n$  increases ( $\Delta^n$  decreases).

We thus have, under the assumption that  $\sigma$  is known, and with the smoothness assumptions on  $F$  and  $\sigma$ , an approximation to the discretized continuous-time likelihood ratio. The probability of false alarm ( $P_{FA}$ ) calculated under this approximation will be exactly that which one would obtain with a discretized version of the exact continuous-time likelihood ratio. The probability of detection ( $P_D$ ) will be an approximation to that which would be obtained using a discretized version of the exact continuous-time likelihood ratio.

In most applications, of course,  $\sigma$  will not be known. We now describe two approaches to obtaining an estimate of  $\sigma$ , corresponding to the two assumptions (A.2) and (A.3), with both based on replacing  $\delta_n[m(x)]$  with  $\underline{m}^n[\underline{x}^n]$  in the expression (5.4) for the discretized log-likelihood ratio.

First, suppose assumption (A.3) is satisfied, and that a training ensemble of representative signal-plus-noise data is available, consisting of  $K$  vectors  $\{\underline{x}^i, i \leq K\}$ , each having  $n$  components, with  $\underline{x}^i(j) = x^i(j\Delta)$ . It is assumed that the vectors represent independent samples. One first applies the matrix  $\underline{F}_n^{-1}$  to each element of this ensemble. Under the assumptions (a) and (b) above, this gives the ensemble of vectors  $\{\underline{\delta Z}^i, i \leq K\}$ , where

$$\begin{aligned} (\underline{\delta Z}^i)(j) &= \underline{Z}^i[(j+1)\Delta^n] - \underline{Z}^i[j\Delta^n] \\ &= \Delta^n \sigma[j, Z^i(j\Delta^n)] + \hat{W}^i[(j+1)\Delta^n] - \hat{W}^i[j\Delta^n]. \end{aligned}$$

One now fixes  $j$ , and uses the  $K$  values having the above expression ( $1 \leq K$ ) to estimate  $\sigma(j, \cdot)$ . Various procedures can be used to carry out the estimation; note that defining  $(\delta \hat{W}^i)(j) \equiv \hat{W}^i[(j+1)\Delta^n] - \hat{W}^i[j\Delta^n]$ , the set  $\{(\delta \hat{W}^i)(j), i \leq K\}$  consists of i.i.d., random variables, with each Gaussian, mean zero, variance  $\Delta^n$ .

Now, with this estimated  $\sigma$  inserted into the expression (5.4) for  $\Lambda^n$ , a sample vector  $\underline{x}^n$  is observed.  $\underline{m}^n[\underline{x}^n]$  is then formed, and used in the expression for  $\Lambda^n$  to form the test statistic  $\Lambda^n[\underline{m}^n[\underline{x}^n]]$ .

If a representative training ensemble of signal-plus-noise data is available, or if  $\sigma$  is known from a mathematical model, then the above procedure gives the preferred mechanization. The algorithm employing this estimate of  $\sigma$  (or using a known  $\sigma$ ) will be denoted as Version II. In the case of non-stationary signal and noise, obtaining an ensemble of S+N data, properly aligned, can be expected to be difficult. However, if the signal-plus-noise is a stationary process, then one may opt to use a long segment of S+N data to estimate a time-invariant  $\sigma$ ; this segment could then be much longer than the observation time over which the detection algorithm is to perform. It can be shown [14] that use of a time-varying  $\sigma$  gives an exact likelihood ratio for the discrete-time problem if the signal-plus-noise process is Gaussian.

Suppose next that nothing is known about the properties of the signal-plus-noise process, and that an ensemble of training data is not available. Version I of the algorithm (5.4) is then implemented as follows, for a fixed value of  $n$ : The observed sample vector  $\underline{x}^n$  is first used to estimate a time-homogenous  $\sigma$ ; the estimated  $\sigma$  is inserted into the expression (5.4) for  $\Lambda^n$ , and then  $\Lambda^n[\underline{m}^n(\underline{x}^n)]$  is evaluated. Thus, this version of the algorithm corresponds to assumption (A.2).

The estimation of  $\sigma$  is made from the single observed sample vector  $\underline{x}^n$  which is to be tested for the presence of a signal. One applies  $\underline{F}_n^{-1}$  to  $\underline{x}^n$ ; assuming that  $\underline{x}^n$  represents signal-plus-noise, this yields a vector  $\underline{\delta Z}$ , which has the representation (under assumptions (a) and (b))

$$(\delta Z)(j+1) = \Delta^n \sigma [Z(j\Delta^n)] + (\delta \hat{W})(j+1).$$

$$(\hat{W})(j+1) = \hat{W}[(j+1)\Delta^n] - \hat{W}[j\Delta^n].$$

The elements of  $\{\hat{W}(j), j \leq K\}$  are i.i.d. random variables, with each Gaussian, mean zero, and variance  $\Delta^n$ .

The preceding discussion will now be summarized. First,  $\sigma$  is either known or else is estimated by one of the two procedures described above when the observation is an  $n$ -component vector  $\underline{x}^n$ . The test statistic, an approximation to the continuous-time log-likelihood ratio under the assumption that the noise has multiplicity  $M = 1$ , then has the expression (with the definition of  $\Lambda^n$  slightly changed from (5.4))

$$\begin{aligned} \Lambda^n(\underline{x}^n) &= \sum_{j=0}^{n-1} \sigma(j, [(\underline{L}_n \underline{F}_n^{-1} \underline{x}^n)_j]) [(\underline{L}_n \underline{F}_n^{-1} \underline{x}^n)_{j+1} - (\underline{L}_n \underline{F}_n^{-1} \underline{x}^n)_j] \\ &\quad - (\Delta^n/2) \sum_{j=0}^{n-1} \sigma^2(j, [(\underline{L}_n \underline{F}_n^{-1} \underline{x}^n)_j]) \\ &= \sum_{j=0}^{n-1} \sigma(j, [(\underline{L}_n \underline{F}_n^{-1} \underline{x}^n)_j]) [(\underline{F}_n^{-1} \underline{x}^n)_{j+1}] - (\Delta^n/2) \sum_{j=0}^{n-1} \sigma^2(j, [(\underline{L}_n \underline{F}_n^{-1} \underline{x}^n)_j]). \end{aligned} \quad (5.6)$$

If now a new data point  $\underline{x}_{n+1}$  is observed, the approximation has the recursive form

$$\begin{aligned} \Lambda^{n+1}(\underline{x}^{n+1}) &= \Lambda^n(\underline{x}) + \sigma[n, (\underline{L}_n \underline{F}_n^{-1} \underline{x})_n] (\underline{F}_{n+1}^{-1} \underline{x})_{n+1} \\ &\quad - (\Delta^n/2) \sigma^2[n, (\underline{L}_n \underline{F}_n^{-1} \underline{x})_n]. \end{aligned} \quad (5.7)$$

The above procedure requires relatively few additional operations when a new data point is observed. The implementation and calculation of  $\Lambda$  require the following operations. First, the function  $\sigma$  must be known and programmed or estimated from the observation. Given the value of  $\Lambda^n(\underline{x}^n)$  and the observation  $\underline{x}^n = (x_1, \dots, x_n)$ , one stores  $\Lambda^n(\underline{x}^n)$ ,  $\underline{x}^n$ ,  $\sigma[n, (\underline{L} \underline{F}_{n-n}^{-1} \underline{x}^n)_n]$ , and  $(\underline{L} \underline{F}_{n-n}^{-1} \underline{x}^n)_n$ . When the data point  $x_{n+1}$  is received, it is only necessary to use the vector  $\underline{x}^{n+1}$  to calculate  $(\underline{F}_{n+1}^{-1} \underline{x}^{n+1})_{n+1}$ , which means to cross-correlate the observation vector  $\underline{x}^{n+1}$  with row  $n+1$  of  $\underline{F}_{n+1}^{-1}$ . This number, say  $b_{n+1}$ , is then used to form  $\Lambda^{n+1}(\underline{x}^{n+1})$ ,

$$\Lambda^{n+1}(\underline{x}^{n+1}) = \Lambda^n(\underline{x}^n) + \sigma[n, \sum_{i=1}^n b_i] b_{n+1} - (\Lambda^n/2) \sigma^2[n, \sum_{i=1}^n b_i] \quad (5.8)$$

Throughout this chapter, we have made the assumption that the noise covariance is known. One then knows  $\{\underline{F}_n, n > 1\}$ , and thus  $\{\underline{F}_n^{-1}, n \geq 1\}$ . As mentioned, each new observation of a data point requires only cross-correlation of the observed vector, an element of  $\mathbb{R}^{n+1}$ , with row  $n+1$  of  $\underline{F}_{n+1}^{-1}$ . It is not necessary to apply the matrix  $\underline{F}_{n+1}^{-1}$  to the data vector.

Implementation of the recursive form of the algorithm is done most conveniently when  $\sigma$  is known, or when a training ensemble of S+N data can be used to estimate  $\sigma$ . If one must estimate  $\sigma$  from the observed data (Version I algorithm), then the recursive formulation given above will need modification. Various approaches can then be used for updating the estimate of  $\sigma$ , depending on the amount of storage available, etc.

The performance of the algorithm can be expected to depend not only on the properties of the data, but also on the sampling rate and the choice of the specific estimation procedure for estimating  $\sigma$ . Thus, implementation for a particular application should be preceded by an extensive study featuring both simulated and experimental data.

## 4. DETECTION IN GAUSSIAN MIXTURE NOISE

### 4.1. Introduction

In this section, the results of the preceding section are extended to nonGaussian noise processes  $(N_t)$  having a representation  $N(\omega, t) = A(\omega)G(\omega, t)$  a.e.  $dP(\omega)dt$  where  $(G_t)$  is a m.s. continuous Gaussian process, vanishing a.s. at  $t = 0$ , and of finite Cramér-Hida multiplicity, and  $A$  is a strictly positive random variable that is independent of  $(G_t)$ . Such noise will be termed a "Gaussian mixture"; if also  $EA^2 < \infty$ , then  $(N_t)$  is spherically-invariant.

As in the preceding section, we will first discuss absolute continuity and representation of the likelihood ratio for the general continuous-time problem. This will be followed by the description of a discrete-time approximation of the continuous-time likelihood ratio, and its implementation.

Although the noise process is a Gaussian mixture, the signal-plus-noise process need not be of this type. Results on absolute continuity and likelihood ratio when both processes are spherically-invariant are given in [25]. [26].

A general analysis of the problems of absolute continuity and likelihood ratio for Gaussian mixture noise, with applications to the Shannon information of communication channels, is contained in [27].

### 4.2. Absolute Continuity

The proofs for the results to be given closely follow those for the case where  $(N_t)$  is Gaussian. In the latter case, the problem of absolute continuity was solved by the following general approach. First, it was shown that versions (and thus their distributions) of the noise and signal-plus-noise processes could be obtained by applying the same measurable transformation (the  $\Phi$  function of Theorem 3.1) to a Gaussian vector martingale

and to a process which resembles a process of "diffusion type" [17]. Then, available results on the Wiener process were exploited to obtain equivalence for the measures induced by the latter pair of processes on  $C_0^M[0,1]$ . These results were then reflected back to the original processes to obtain the results on absolute continuity. The existence of the measurable transformation with the properties given in Theorem 3.1 does not require that  $(N_t)$  be Gaussian. Thus, to extend the results on absolute continuity from the Gaussian  $(N_t)$  to that of a Gaussian mixture  $(N_t)$ , one must only show that the martingale results used in the Gaussian case can be extended to the Gaussian mixture problem. The Gaussian martingale results rely on Girsanov's theorem, which can be summarized as follows. If  $(M_t)$  is a continuous local martingale, with natural increasing process  $(\langle M \rangle_t)$ , and if  $(f_t)$  is a predictable process, then  $(\tilde{M}_t)$  defined by

$$\tilde{M}_t = \int_0^t f(s) d\langle M \rangle_s + M_t$$

is a continuous local martingale with  $(\langle M \rangle_t)$  as its natural increasing process, with respect to the measure  $Q$  defined by

$$\frac{dQ}{dP} = \exp\left\{-\int_0^t f dM - \frac{1}{2} \int_0^t f^2 d\langle M \rangle\right\},$$

and assuming that  $Q$  is a probability.

When  $(M_t)$  is a Gaussian martingale, its continuity and the form of  $(\langle M \rangle_t)$  suffice to completely specify the law induced by  $(M_t)$ ; it is a transformation of Wiener measure. This is no longer true when  $(\langle M \rangle_t)$  is random. In fact, the identification of the law of  $(M_t)$  with the knowledge of  $(\langle M \rangle_t)$  is a difficult unsolved problem [28]. In the case of Gaussian-mixture noise  $(\underline{A}\underline{B}_t)$ , where  $(\underline{B}_t)$  is a vector Gaussian martingale, then  $\langle \underline{A}\underline{B} \rangle = A^2 \underline{\Gamma}_{\underline{B}}$ , where  $\underline{\Gamma}_{\underline{B}}$  is the covariance matrix of  $\underline{B}$ , and it can be shown [27] that  $\langle \underline{A}\underline{B} \rangle$  determines the law

of  $\underline{A}\underline{B}$ . This identification is done in two steps. First, one shows that if a continuous local martingale  $\underline{M}$  is such that  $\langle \underline{M} \rangle = A^2 \Gamma_{\underline{M}}$ ,  $\Gamma_{\underline{M}}$  a covariance matrix and  $A$  a strictly-positive random variable, then  $\underline{M} = \underline{A}\underline{B}$  where  $\underline{B}$  is a Gaussian martingale independent of  $A$ . Secondly, one shows that the law of  $A$  with respect to  $Q$  (defined above) is the same as the law of  $A$  with respect to  $P$ .

Prop. 4.1. If  $\underline{B} \equiv \underline{M}/A$ , then  $\underline{B}$  is a Gaussian martingale independent of  $A$ .

This result is clear; independence of  $A$  and  $\underline{B}$  follows from the fact that  $A$  is measurable w.r.t. each  $\sigma$ -field contained in the underlying filtration.

Prop. 4.2. If  $U$  is a bounded function measurable w.r.t. all the  $\sigma$ -fields of the underlying filtration, then  $E_Q U = E_P U$ .

This result follows from  $E_Q U = E_P U D_t$ , where  $(D_t)$  is the martingale obtained by taking the conditional expectation of the Radon-Nikodym derivative  $D \equiv dQ/dP$ .  $\lim_{t \downarrow 0} D_t = 1$  in  $L_1[dP]$ ; one successively applies this equality to the characteristic functions of  $A$  and  $\underline{M}$ .

The results to be given now are based on martingale arguments that rely on Prop. 4.1 and Prop. 4.2. The next result is the generalization of Theorem 3.3.

Theorem 4.1. Suppose that  $(Y_t)$  is a stochastic process adapted to  $\sigma(A) \vee \underline{\sigma}(G) \vee \underline{\sigma}(V)$ , where  $(V_t)$  is any process independent of  $(G_t)$  and  $A$ .

- (1) If  $Y(t) = S(t) + N(t)$  a.e.  $dP$  for each fixed  $t$ , where  $(S_t)$  is a stochastic process adapted to  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$  and with almost all paths in  $H_N$ , then  $\nu_Y \ll \nu_N$ .
- (2) If (1) holds, and also  $Y(t) = S(t) + N(t)$  a.e.  $dPdt$ , then  $\mu_Y \ll \mu_N$ .

As can be seen, the above sufficient conditions for existence of the likelihood ratio are very similar to those of Theorem 3.3. Similarly, the representations of the likelihood ratio when these conditions are satisfied are identical to that given in Theorem 3.7. Those representations are given in [27]. In order to completely describe the likelihood ratio, it is necessary to give the likelihood ratio  $dP_{\underline{Z}}/dP_{AB}$ , where  $(\underline{B}_t)$  is a Gaussian vector martingale,  $A$  is a strictly positive random variable independent of  $(\underline{B}_t)$ , and  $(\underline{Z}_t)$  is a process with paths in  $C_0^M[0,1]$  and such that  $P_{\underline{Z}} \ll P_{\underline{B}}$ . The likelihood ratio  $dP_{\underline{Z}}/dP_{AB}$  is given in [27], along with the appropriate definitions. The derivation and consequent statement of that general result is technically somewhat involved, and will be omitted. Instead, we make a stronger set of assumptions that will suffice for eventual implementation, similar to the results given in [16].

Before giving the result on likelihood ratios under these stronger assumptions, we state an important result that is essential to eventual implementation of the approximated likelihood ratio. In Section 3.3, likelihood ratios (Theorem 3.7) were stated in terms of a function  $\underline{m}$  having range in  $C_0^M[0,1]$  and domain in either  $\mathbb{R}^{[0,1]}$  or  $L_2[0,1]$ . The same functions will occur in the Gaussian mixture problem, but now one must consider the presence of the random variable  $A$ . This is simplified by the following result.

Prop. 4.3. Let  $\underline{m}_a: \mathbb{R}^{[0,1]} \rightarrow C_0^M[0,1]$  be the  $\underline{m}$  defined in Theorem 3.6, when  $A = a$ . Let  $\underline{m}'_a: L_2[0,1] \rightarrow C_0^M[0,1]$  be the function  $\underline{m}$  defined in (2) of Theorem 3.7, when  $A = a$ . Then  $\underline{m}_a$  and  $\underline{m}'_a$  are independent of the value of  $a$ .

With Prop. 4.3, likelihood ratios can be obtained. The following result



is the equivalent of Theorem 3.7.

Theorem 4.2. Suppose that  $(V_t)$  is a stochastic process independent of  $(N_t)$ , and that  $(S_t)$  is a stochastic process adapted to  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$  with almost all paths in  $H_G$ . Then  $S_t = \sum_{i=1}^M \int_0^t F_i(t,s) \phi_i(s) d\beta_i(s)$ , where  $(\phi_i(s))$  is  $\underline{\sigma}(N) \vee \underline{\sigma}(V)$ -predictable and has paths a.s. in  $\mathcal{L}_2[\beta_i]$ . Define  $(Z_t)$  to be the  $C_0^M$  process with  $i$ th component given by  $Z_i(t, \omega) = \int_0^t \phi_i(s, \omega) d\beta_i(s) + AB_i(t, \omega)$ .  $P_{\underline{Z}} \ll P_{AB}$  and  $S_t + N_t = \sum_{i=1}^M \int_0^t F_i(t,s) dZ_i(s)$ .  
Then

(1) Suppose  $Y_t = S_t + N_t$  a.e.  $dP$ , for each fixed  $t$  in  $[0,1]$ . Then  $[dv_Y/dv_N](y) = [dP_{\underline{Z}}/dP_{AB}](\underline{m}y)$  a.e.  $dv_N(y)$ , where  $\underline{m}$  is defined as in Theorem 3.6.

(2) Suppose  $Y_t = S_t + N_t$  a.e.  $dt dP$ . Then  $[d\mu_Y/d\mu_N](x) = [dP_{\underline{Z}}/dP_{AB}](\underline{m}x)$  a.e.  $d\mu_N(x)$ , where  $\underline{m}(y)$  in  $C_0^M$  has as its  $i$ th component  $[\underline{m}(y)]_i(t) = \sum_{n \geq 1} \langle y, e_n \rangle \langle f_t^i, e_n \rangle / \lambda_n$ .

(3) Suppose that  $Y_t = S_t + N_t$  a.e.  $dP$  for each  $t$  in  $[0,1]$ , and that  $(Y_t)$  has continuous paths. Then  $[d\lambda_Y/d\lambda_N](x) = [dP_{\underline{Z}}/dP_{AB}](\underline{m}i_0 x)$  a.e.  $d\lambda_N(x)$ , where  $\underline{m}$  is defined in (2) and  $i_0: C_0^M[0,1] \rightarrow L_2[0,1]$  is the injection map.

The following result is a special case of Theorem 4.1 and Theorem 4.2.

Theorem 4.3. Let  $(V_t)$  be any process independent of  $(N_t)$ . Suppose that  $(Y_t)$  is a stochastic process having the representation  $Y_t = S_t + N_t$  a.e.  $dP dt$  where  $(S_t)$  is adapted to  $\underline{\sigma}^\circ(N) \vee \underline{\sigma}^\circ(V)$ . For fixed  $a$ , let  $(S_t^a)$  be the process  $(S_t)$  when  $\Lambda$  is fixed and  $\Lambda = a$ , and set  $Y_t^a \equiv S_t^a + aG_t$ . Suppose that for each fixed  $a$  in  $\text{range}(\Lambda)$ ,  $(S_t^a)$  has almost all paths in  $H_G$ . Finally, suppose that  $\Lambda$  is a discrete random variable with range in

$\{a_i, i \geq 1\}$ . Then

(a)  $v_Y \ll v_N$  and  $v_{Y^a} \ll v_{aG}$  for every  $a$  in  $\{a_i, i \geq 1\}$ ;

$$(b) \quad \frac{dv_Y}{dv_N}(x) = \sum_{i \geq 1} I_{C(a_i)}(x) \left[ \frac{dv_{S^a_i + a_i G}}{dv_{a_i G}} \right] \quad \text{a.e. } dv_N(x).$$

where  $I$  denotes the indicator function and  $C(a_i) \equiv$

$$\{x: \overline{\lim}_n \frac{1}{n} \sum_{j=1}^n \frac{\langle x, e_j \rangle^2}{\lambda_j} = a_i^2\}.$$

Using Theorem 4.3 and Prop. 4.3, one can apply the results previously obtained for Gaussian noise to obtain the likelihood ratio in the Gaussian mixture case. Theorem 4.3 contains two basic assumptions that considerably simplify the general result. One is the assumption on smoothness of the process  $(S_t^a)$  for all  $a$  in the support of  $P \circ A^{-1}$ . The second is the assumption (to give the form of the likelihood ratio) that  $A$  has discrete support. Neither is necessary, and neither is made in [27], where the general results are obtained. However, they are reasonable for applications, which is the eventual aim here. Note that under the assumptions of Theorem 4.3,

$$A^2(\omega) = \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \langle x(\omega), e_j \rangle^2 / \lambda_j$$

with probability one under both  $P_N$  and  $P_{S+N}$ , by the law of large numbers; in the case of  $P_{S+N}$ , one uses the fact that  $(S_t)$  has paths a.s. in  $H_N$ . This observation, Theorem 4.3, and Prop. 4.3 permit one to specify a likelihood ratio for implementation quite naturally, as seen in the following section.

#### 4.3. Approximation and Implementation of the Likelihood Ratio

Here we shall make assumptions similar to (A.1), (A.2) and (A.3) of

Section 3.4. We also assume  $EA^2 = 1$ , so that  $R_N = R_G$ , and that  $A$  is a discrete-random variable.

Version I of the algorithm is based on assumptions A.1 and A.2 below.

(A.1) The Gaussian process  $(G_t)$  defining the Gaussian mixture noise process has multiplicity  $M = 1$ , and the process  $(B_1(t))$  is the standard Wiener process,  $(W_t)$ . Thus

$$N(t) = AG(t) = A \int_0^t F(t,s) dW(s),$$

with  $F$  a Volterra kernel such that  $\int_0^1 \int_0^1 F^2(t,s) ds dt < \infty$ .

(A.2) The signal-plus-noise process  $(Y_t)$  is given by

$$Y(t) = \int_0^t F(t,s) dZ(s),$$

where  $(Z_t)$  has the representation

$$Z(t) = \int_0^t \theta[Z(s)] ds + \hat{A}\hat{W}(t)$$

a.e. dP for each  $t$ ; and  $P\{\omega: \int_0^1 \theta^2[Z(\omega,s)] ds < \infty\} = 1$ .

Version II is based on A.1 and on A.3, below.

(A.3)  $Y(t)$  has the representation given in (A.2), but now

$$Z(t) = \int_0^t \sigma[s, Z(s)] ds + \hat{A}\hat{W}(t)$$

where  $P\{\omega: \int_0^1 \sigma^2[s, Z(\omega,s)] ds < \infty\} = 1$ .

Implementations of the discrete-time approximation to this likelihood ratio are obtained from those given in Section 3.4 for the Gaussian noise case, as follows. Here we assume the notation of Section 3.4:  $R_N = \Delta^n F_{-n}^* F_n$ , and  $L_n$  is the summation matrix.  $\Delta^n$  is the (uniform) sampling interval, and we assume an observation in  $\mathbb{R}^n$ . We first discuss the implementation of the

algorithm with time-homogenous drift, with no prior knowledge of the signal-plus-noise properties (Version I).

Step 1. Form the vector  $F_n^{-1} \underline{x}^n$ . This gives, if signal is present and with the smoothness assumptions (a) and (b) of Section 3.5, the vector  $(\delta Z^n)$  where

$$(\delta Z^n)(i+1) \cong \Delta^n \sigma[Z(i\Delta^n)] + a\hat{W}([i+1]\Delta^n) - a\hat{W}(i\Delta^n).$$

Step 2. Estimate  $a$ ; for example, by computing the sample quadratic variation of  $Z^n$ ; that is,  $(\hat{a})^2 \equiv \sum_{i=1}^{n-1} [(\delta Z^n)(i)]^2$ .

Step 3. Assuming that  $(\delta Z^n)(i+1) = \Delta^n \sigma[Z(i\Delta^n)] + \hat{a}[\hat{W}([i+1]\Delta^n) - \hat{W}(i\Delta^n)]$  for  $i \leq n-1$ , estimate  $\sigma$ .

Step 4. Insert  $\hat{\sigma}$  into the expression (5.4) of Section 3, for  $\Lambda^n$ , and then evaluate  $\Lambda^n[\underline{m}^n[\underline{x}^n]]/(\hat{a})^2$ .

Note that in this Version I of the algorithm, the estimate of  $\sigma$  may depend on the estimate of  $a$ . It should also be noted that the above estimation procedure for  $a$  assumes large  $n$  and small  $\Delta^n$ .

In the likelihood ratio with time-varying  $\sigma$ , as in assumption (A.3), one can proceed by first estimating  $\Lambda$  for each sample path of the training ensemble, and then estimating  $\sigma(i, \cdot)$  for each  $i \leq n$  by using these estimated values of  $\Lambda$ . Of course, a better solution would be to know  $\sigma$  from a mathematical model, if such a model were available.

The algorithms given in Section 3.4 are special cases of these algorithms, obtained by setting  $\hat{a} \approx 1$ .

## 5. COMMENTS ON ROBUSTNESS AND RANGE OF APPLICATIONS OF THE DETECTION ALGORITHMS

The results summarized in this chapter, and given in detail in [17] and [27], give a complete set of sufficient conditions and necessary conditions for the existence of a likelihood ratio, under the assumption that the noise process  $(N_t)$  is a Gaussian mixture process with the following properties.

- (1)  $(N_t)$  has a representation  $(AG_t)$ , where  $(G_t)$  is Gaussian,  $A$  is a positive random variable independent of  $(G_t)$ , and  $(G_t)$  is m.s. continuous.
- (2)  $N(0) = 0$  with probability one.
- (3)  $(G_t)$  has finite Cramér-Hida multiplicity.

These assumptions are very weak. (1) is typically satisfied in applications. (2) can be circumvented in applications by assuming that the problem started at a time instant prior to the first sampling time. (3) is practically meaningless, since the multiplicity is permitted to be arbitrarily large.

The adjective "complete" in describing the above-mentioned results should be interpreted as follows. The set of results obtained for this problem are exactly those that have been known for many years when it is assumed that the noise is the Wiener process. This is all that one can expect; the unusual properties of the Wiener process, previously mentioned, permit one to obtain sufficient conditions and necessary conditions for absolute continuity when the noise is the Wiener process. Noise processes encountered in applications typically do not have the fortuitous properties of the Wiener process. However, through use of the spectral (Cramér-Hida) representation of second-order stochastic processes, the Girsanov theorem and extensions, and various results from stochastic calculus, it has been possible to obtain exactly comparable results for the class of Gaussian and nonGaussian noise processes

described above.

Under the same assumptions, expressions for the likelihood ratio have been obtained when the conditions sufficient for its existence are satisfied.

These general results, however, are only mathematical results without further work. That is, they must be given an interpretation such that they can be implemented without requiring unrealistic knowledge of the parameters appearing in the expressions for the likelihood ratio.

Implementation of discrete-time approximations to the exact continuous-time likelihood ratio would require knowledge of the spectral (Cramér-Hida) representation of the noise. The representation will typically not be known; its determination is well-known to be a very difficult unsolved problem (for processes of a general type). Thus, additional assumptions are necessary to permit implementation of the likelihood ratios as signal detection algorithms. The basic additional assumptions are that the noise process is of unit spectral multiplicity, and that the signal-plus-noise process is a filtered diffusion. The first of these two assumptions can be justified from the fact that unit-multiplicity second-order processes are dense (in a mean-square sense) in the linear space of all mean-square continuous processes and that all wide-sense stationary processes and all discrete-time second-order processes are of unit multiplicity, and the second still permits one to consider a very large class of signal-plus-noise processes under the assumption that a likelihood ratio exists.

With these additional assumptions, two discrete-time approximations to the likelihood ratio have been given. These approximations have very reasonable implementations, requiring prior knowledge of the noise covariance matrix. In the fully adaptive version, the remaining parameter of the detection algorithm can be easily obtained. For Version II, if a good

mathematical model is not available, then this parameter is best obtained from a representative sample of "training" signal-plus-noise data.

The amount of training data required may be less than that required to obtain a reliable estimate of the S+N covariance matrix; thus, the Version II algorithm can be used in some situations where the S+N process is Gaussian but the S+N covariance cannot be reliably determined. In many applications, as for example the highly-complicated world of underwater acoustics, mathematical-physical models of important signal-plus-noise processes have been sought for many years, generally with only limited success. Even when such models are derived, they may be the result of a great many approximations, and may require detailed knowledge that is not typically available. It is clearly desirable to have detection algorithms that do not rely on such models for their effectiveness. From this aspect, the fully adaptive Version I algorithm is preferable.

Of course, under the assumptions used here, implementation of the detection algorithms requires only knowledge of the drift of a diffusion. For a particular application, a successful modeling effort resulting in determination of the drift function would enable the implementation of the more powerful Version II algorithm without the need to have an ensemble of training data.

Numerical evaluations of the two algorithms, using both simulated and experimental data, have given encouraging results.

As can be seen from the preceding development, the theoretical basis of these two algorithms and their implementable form give reasons to believe that they may provide, at least in many applications, a satisfactory solution to a long-standing and much-encountered detection problem: detection of Gaussian or nonGaussian signals imbedded in Gaussian or Gaussian mixture noise.

## 6. CONCLUDING REMARKS.

Many discrete-time finite-sample detection algorithms are obtained from consideration of only the discrete-time (and finite-dimensional) problem. If this is done, and the data represents discretized continuous-time data, then the problem of developing an optimally-effective algorithm is akin to that which the blind man faces in describing the elephant. It is obviously preferable, if possible, to develop a discrete-time algorithm based on approximations to the likelihood ratio for the continuous-time problem.

However, likelihood ratios for continuous-time problems are inextricably bound (if done correctly) to the problem of existence of the likelihood ratio, or absolute continuity. Unfortunately, studies on existence of likelihood ratios are considered to be of only mathematical interest by many engineers concerned with systems design and by many applied statisticians concerned with inference. This attitude may well be justified if the studies are on models not suitable for applications, or if the results of the study are arcane theorems left in a form unintelligible to potential users, or if the results are expressions that require knowledge of quantities that one cannot realistically expect to be known. As can be seen from the preceding development, appropriate studies on absolute continuity and likelihood ratios for continuous-time problems can be extremely important in developing practical discrete-time detection algorithms.

## 7. ACKNOWLEDGEMENTS

The mathematical research summarized here was supported by grants from the Mathematical Sciences Division of the Office of Naval Research and by the Networking and Communications Research and Infrastructure Program of the National Science Foundation. ONR-MSD also funded the establishment of a



dedicated computer facility; an equipment grant from the Mathematical and Information Sciences Division of the Air Force Office of Scientific Research provided a major upgrade. Computational work on data analysis, modeling, and algorithm evaluations was supported by the Signal Processing Branch of the Naval Oceans Systems Center. We thank each of these organizations, and especially the cognizant program directors, for their support.

## REFERENCES

- [1] C.W. Helstrom. *Statistical Theory of Signal Detection*, Pergamon, New York (1960).
- [2] H.L. van Trees, *Detection, Estimation, and Linear Modulation Theory, Part I*, Wiley, New York (1968).
- [3] C.R. Rao and V.S. Varadarajan. Discrimination of Gaussian processes, *Sankhya Ser. A*, 25, 303-330 (1963).
- [4] M. Rosenblatt, ed. *Proceedings of the Symposium on Time Series Analysis*, Wiley, New York (1963). See Chapter 11 (E. Parzen), Chapter 19 (G. Kallianpur and H. Oodaira), Chapter 20 (W.L. Root), and Chapter 22 (A.M. Yaglom).
- [5] T.E. Duncan. Likelihood functions for stochastic signals in white noise, *Inf. Control*, 16, 303-310 (1970).
- [6] T.T. Kadota and L.A. Shepp. Conditions for absolute continuity between a certain pair of probability measures, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 16, 250-260 (1970).
- [7] T. Kailath and M. Zakai. Absolute continuity and Radon-Nikodym derivatives for certain measures relative to Wiener measure, *Ann. Math. Statist.*, 42, 130-140 (1971).
- [8] J. Memin. Sur quelques problèmes fondamentaux de la théorie du filtrage, *These (troisième cycle)*, U.E.R. Mathématiques et Informatique, University of Rennes (1974).
- [9] R.S. Liptser and A.N. Shiryaev. *Statistics of Random Processes, I. General Theory*, Springer, Berlin-Heidelberg-New York (1977).
- [10] C.R. Baker. On equivalence of probability measures, *Annals of Probabil.*, 1, 690-698 (1973).
- [11] J.L. Lawson and G.E. Uhlenbeck. *Threshold Signals*, McGraw-Hill, New York, pp. 161-163 (1950).
- [12] C.R. Baker. Optimum quadratic detection of a random vector in Gaussian noise, *IEEE Trans. Comm. Tech.*, 14, 802-805 (1966).
- [13] C.R. Baker. On the deflection of a quadratic-linear test statistic, *IEEE Trans. Inform. Theory*, 15, 16-21 (1969).
- [14] C.R. Baker. Absolute continuity of measures on infinite-dimensional spaces, *Encyclopedia of Statistical Sciences*, Vol. 1, 3-11, John Wiley & Sons, New York (1982).
- [15] C.R. Baker and R.E. Cunningham. Results of statistical tests on active sonar data, *USN J. Underwater Acoustics*, 26, 247-264 (1976).

- [16] C.R. Baker and A.F. Gualtierotti. Likelihood ratios and signal detection for nonGaussian processes, *Stochastic Processes in Underwater Acoustics*, Lecture Notes in Control and Information Science, 85, 154-180, Springer-Verlag, Berlin (1986).
- [17] C.R. Baker and A.F. Gualtierotti. Discrimination with respect to a Gaussian process, *Probabil. Theory Rel. Fields*, 71, 159-182 (1986).
- [18] H. Cramér. On some classes of non-stationary stochastic processes, *Fourth Berkeley Sympos. Math. Stat. and Probabil.*, 2, 57-77, Univ. Calif. (1961).
- [19] T. Hida. Canonical representations of Gaussian processes and their representations, *Mem. Coll. Science*, 33A, 109-115, Univ. Kyoto (1960).
- [20] K. Itô, The topological support of Gaussian measure on Hilbert space, *Nagoya Math. J.* 38, 181-183.
- [21] J.L. Doob. *Stochastic Processes*, Wiley, New York (1950).
- [22] K.R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic press, New York (1967).
- [23] D. Freedman. *Markov Chains*, Springer, Berlin-Heidelberg-New York (1983).
- [24] A. Ephremides, A property of random processes with unit multiplicity, *J. Multivariate Analysis*, 7, 525-534 (1977).
- [25] A.F. Gualtierotti. Some remarks on spherically-invariant distributions, *J. Multivariate Anal.*, 4, 347-349 (1974).
- [26] A.F. Gualtierotti, A likelihood ratio formula for spherically-invariant processes, *IEEE Trans. Inform. Theory*, IT-22, 610 (1976).
- [27] C.R. Baker and A.F. Gualtierotti. Absolute continuity and mutual information for Gaussian mixtures, LISS-28, Department of Statistics, University of North Carolina, Chapel Hill (1988).
- [28] J. Jacod. *Calcul Stochastique et Problemes de Martingales*, Lecture Notes in Mathematics, 714, Springer, Berlin (1979).